Guochang Xu
(Ed.)

Sciences of Geodesy-I

Advances and Future Directions

Springer

Sciences of Geodesy – I

Guochang Xu
Editor

Sciences of Geodesy – I

Advances and Future Directions

Springer

Editor
Dr. Guochang Xu
GFZ German Research Centre for Geosciences
Department 1: Geodesy and Remote Sensing
Telegrafenberg
14473 Potsdam
Germany
xu@gfz-potsdam.de

ISBN 978-3-642-11740-4 e-ISBN 978-3-642-11741-1
DOI 10.1007/978-3-642-11741-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010929615

Cover illustration: Courtesy U.S. National Oceanic and Atmospheric Administration (NOAA).

Cover design: deblik, Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This series of reference books describes sciences of different fields in and around geodesy with independent chapters. Each chapter covers an individual field and describes the history, theory, objective, technology, development, highlights of research and applications. In addition, problems as well as future directions are discussed. The subjects of this reference book include Absolute and Relative Gravimetry, Adaptively Robust Kalman Filters with Applications in Navigation, Airborne Gravity Field Determination, Analytic Orbit Theory, Deformation and Tectonics, Earth Rotation, Equivalence of GPS Algorithms and its Inference, Marine Geodesy, Satellite Laser Ranging, Superconducting Gravimetry and Synthetic Aperture Radar Interferometry. These are individual subjects in and around geodesy and are for the first time combined in a unique book which may be used for teaching or for learning basic principles of many subjects related to geodesy. The material is suitable to provide a general overview of geodetic sciences for high-level geodetic researchers, educators as well as engineers and students. Some of the chapters are written to fill literature blanks of the related areas. Most chapters are written by well-known scientists throughout the world in the related areas.

The chapters are ordered by their titles. Summaries of the individual chapters and introductions of their authors and co-authors are as follows.

Chapter 1 "Absolute and Relative Gravimetry" provides an overview of the gravimetric methods to determine most accurately the gravity acceleration at given locations. The combination of relative and absolute gravimeters allows the surveying of local, regional and global networks which can be used to monitor short-term and long-term gravity variations. As an example of the present state-of-the-art absolute and relative gravimeters, the main characteristics and accuracy estimates for the Hannover instruments are presented. The observational g-values are reduced for the time-dependent and position-dependent gravity variations due to Earth's body and ocean tides, atmospheric mass redistributions and polar motion. Usually hydrological effects are not reduced but they may become a target signal to monitor changes in aquifers and deep water reservoirs. The gravimetric surveying of the crustal deformation in northern Europe is still a main focus of the ongoing absolute gravimetry activities. It serves to study the postglacial isostatic adjustment of Fennoscandia.

The author of Chap. 1 is Dr. Ludger Timmen. Ludger Timmen works at the Leibniz Universität Hannover (LUH), Germany, where he has lectured in gravimetry since 2005 (since 1996 as a guest lecturer). He holds a Dipl.-Ing. degree in surveying engineering and obtained a Ph.D. from the University of Hannover (now LUH) in 1994. As a research assistant at the Institut für Erdmessung (IfE) of LUH, he specialised in precise gravimetry and its application to geodynamic research (tectonics, Earth tides). From 1995 to 1999, he held a scientist position at GFZ Potsdam, the German geoscience research center, focussing on airborne gravimetric techniques and coordinating the international airborne gravimetry projects of GFZ. Back at IfE since 2002, his main research interest is the improvement and application of relative and absolute gravimetry to measure small temporal gravity variations on the timescale from some days to a few decades. He organized and performed various gravimetry campaigns in China, South America and northern Europe and participated in two German Antarctic expeditions.

In Chap. 2 "Adaptively Robust Kalman Filters with Applications in Navigation", the main achievements of the adaptively robust filter are summarized from the published papers in recent years. In Sect. 2.1, the background and developments of adaptive filters are summarized. The principle of the adaptively robust filter is presented and the estimators are derived in Sect. 2.2. The special cases of the new adaptively filter are also given. In Sect. 2.3, the properties of the adaptive Kalman filter are analysed. After that the establishment of four kinds of learning statistics for judging the kinematic model errors, which include state discrepancy statistic, predicted residual statistic, variance component ratio statistic and velocity discrepancy statistic are given in Sect. 2.4. And in Sect. 2.5, four adaptive factors for balancing the contribution of kinematic model information and measurements are presented, which include three-segment function, two-segment function, exponential function and zero and one function for state component adaptation. In Sect. 2.6, two fading filters and adaptively robust filter are compared and computation examples are included. In Sect. 2.7, the Sage adaptive filter and an adaptively robust filter are also compared; the problems of the Sage adaptive filter are analysed. The last section presents some application examples of the adaptively robust filter.

The author of Chap. 2 is Prof. Yuanxi Yang. Yuanxi Yang, Academician of Chinese Academy of Science, is a professor. He graduated in Geodesy in 1980 and 1987 from the Zhengzhou Institute of Surveying and Mapping (ZISM) with BSc and MSc degrees. He obtained his doctorate from Institute of Geodesy and Geophysics, the Chinese Academy of Science, in 1991. He worked as associate professor and professor in ZISM from 1990 to 1992 and from 1992 to 1998, respectively. He has been a deputy director and chief engineer of Xi'an RISM since 1998. He was a visiting scholar of Center for Space Research of University of Texas, USA in 1995. From 1996 to 1997 he was a scientist in Institute of Theoretical Geodesy of Bonn University in Germany under a Humboldt fellowship. He is a member of Chinese Union of Geodesy and Geophysics since 1997, second secretary of Section IV, IAG from 1999 to 2003, and member of ICCT, IAG since 1999. His main research field includes geodetic data processing, navigation and geodetic coordinate system, etc.

He has published more than 100 papers on robust estimation and adaptive Kalman filtering.

Chapter 3 "Airborne Gravity Field Determination" outlines some of the basic principles of airborne gravimetry, with special focus on *geodetic* applications, and gives some examples from recent large-scale surveys. For geodesy, the main focus is more on absolute accuracy and long-wavelength stability, since long-wavelength errors in gravity transforms to large geoid errors; for geophysical exploration focus is mainly on the short-wavelength performance and ultimately making reliable detection and mapping of small, elusive gravity signatures. The chapter starts with an introduction and describes principles of airborne gravimetry and filtering technique of airborne gravity. Some results of large-scale government airborne surveys are given in Sect. 3.4 and downward continuation of airborne gravimetry are discussed in Sect. 3.5. Geoid determination and conclusions are given in the sixth section and the last, respectively.

The author and co-author of Chap. 3 are state scientist Rene Forsberg and Dr. Arne V. Olesen.

Rene Forsberg is the state geodesist and head of the Department of Geodynamics of the National Space Institute of Denmark, formerly known as the Danish National Space Center. He obtained MSc degrees in both geophysics and geodesy from University of Copenhagen during 1980s before joining the Danish Geodetic Institute as research geodesist with working fields as gravimetry, satellite geodesy and Greenland survey projects. From 1983 to 1984 he was a visiting scientist of Ohio State University and University of Calgary (1984–1985). Rene Forsberg has been an external lecturer at University Copenhagen since 1989. He is a project coordinator or participant in numerous ESA, EU and research council projects, focusing on gravity field determination or cryosphere measurement. In addition, he is a member of the scientific advisory board for the ESA Cryosat mission, chairman of the IAG International Gravity Field Service, the vice president of the International Gravity and Geoid Commission and a member of the International Association of Geodesy Restructuring Committee since 1999. He was elected as the chairman of IAG Special Working Group "Local Gravity Field Modelling" (1987–1995) and appointed as section president (Gravity field Determination) of IAG (1995–1999). Rene Forsberg is a world-renowned scientist in the field of aerogravimetry. Several PhD studies were completed under his supervision and he is the author and co-author of more than 250 scientific papers in journals, proceedings and reports.

Arne V. Olesen is a senior scientist in National Space Institute of Denmark. He obtained his doctorate 2001 in University of Copenhagen and worked as scientist in National Survey and Cadastre, Denmark, since 1997. He has been working very intensively on aerogravimetry research and field campaigns as well as GPS investigation since many years and authored and co-authored many scientific papers.

The Chap. 4 "Analytic Orbit Theory" describes the satellite orbit theory in a condensed way. The perturbed equations of satellite motion are discussed first. Then singularity-free and simplified equations are given. The solutions of extraterrestrial

disturbances such as solar radiation pressure, atmospheric drag and the disturbance of the sun, the moon and planets are then outlined. Solutions of geopotential disturbances are given with examples. Numerical and analytical orbit determination is dealt with before the summary and discussions.

I (Guochang Xu) am the author of Chap. 4. After graduating in mathematics and geodesy from Wuhan University and the Chinese Academy of Sciences (CAS) in 1982 and 1984, respectively, I obtained Dr.-Ing. degree from the Technical University (TU) Berlin in 1992. Having worked as a research associate at the TU Berlin from 1986 to 1993, as a scientist at the GeoForschungsZentrum (GFZ) Potsdam from 1993 to 1998 and as a senior scientist at the National Survey and Cadastre, Denmark, from 1998 to 1999, I returned to the GFZ as a senior scientist in 1999. I have been involved in geodetic research since 1983 and have authored and co-authored several scientific books and software. From 2003 to 2008 I was an overseas assessor, adjunct professor and winner of overseas outstanding scholar fund of CAS. I am an adjunct professor of ChangAn University since 2003, overseas communication assessor of Education Ministry China since 2005, an adjunct professor of National Time Service Center, CAS, since 2009 and national distinguished expert of China Academy of Space Technology since 2010.

The Chap. 5 "Deformation and Tectonics" addresses some aspects of the use of the GPS system in the study of plate tectonics. After a short summary on the evolution of models of the angular velocities of plate tectonics using geophysical, geological and geodetic data, the best methodologies to define a reference frame using GPS base stations are explained and the problem of mapping the GPS solutions to accurately obtain the position of a station with respect to the most recent International Terrestrial Reference Frame solution (ITRFxxxx) is discussed in Sect. 5.3. In the next section, the geophysical signals that need to be subtracted from the GPS observations to clearly distinguish the secular tectonic plate motion are referred. In Sect. 5.5 the problem of estimating the plate motion using those preprocessed GPS time-series is described. The contribution of the GPS technique to unravel the geodynamics features of a plate boundary zone is exemplified using research carried out in the Azores Triple Junction region. The importance of a full integration of all available GPS data, both continuous and episodic, possible evolutions in the exploitation of the GNSS technology, including the benefits of a multi-technique approach, as well as the need for a proper integration of geodetic, geophysical and geological information are stressed in the last section.

The author and co-authors of Chap. 5 are Dr. Luisa Bastos, Dr. Machiel Bos and Dr. Rui Manuel Fernandes.

Luisa Bastos is a senior researcher at the University of Porto and since 1997 director of the Astronomical Observatory of the Faculty of Sciences. Since 2002 she is a member of CIIMAR (Centre of Marine and Environmental Research of the University of Porto). She graduated as surveying engineer in 1976 at the University of Porto where she received a Ph.D. degree in 1991 with a work focused on GPS application to geodynamics. Her main interest is on precise GNSS applications and

in the last 20 years she has been involved not only in projects related with geodynamics studies, but also in the development of applications based in the integration of GNSS with other sensors and its exploitation for airborne, terrestrial and marine applications, namely airborne gravimetry and mobile mapping. She has been supervising or co-supervising M.Sc. and Ph.D. thesis on these topics. From 1999 to 2004 she acted as president of the WEGENER project and is presently a member of the WEGENER inter-commission. She is currently working on research projects that involve the exploitation of satellite-based systems and multi-sensor integration for geodynamics studies, environmental monitoring and coastal dynamics.

Machiel Bos studied aerospace engineering at Delft University of Technology, The Netherlands. After his graduation in 1996 he performed his Ph.D. research at Proudman Oceanographic Laboratory, Liverpool, United Kingdom. In 2001 he spent 7 months as post-doc at Onsala Space Observatory, Sweden. From 2001 to 2003 he worked as a post-doc at the Faculty of Geodesy of Delft University of Technology. From 2003 to 2008 he held a post-doc position at the Astronomical Observatory of Porto, Portugal, and he is working at CIIMAR since 2008 (Centre of Marine and Environmental Research of the University of Porto). His main scientific interests are ocean tide loading, GPS time-series analysis and the geoid.

Rui Manuel Fernandes has a doctoral degree in earth and space sciences from Technical University of Delft (The Netherlands). He is assistant professor in the University of Beira Interior (UBI), Covilhã, Portugal, and associated researcher of Institute Geophysical Infante D. Luíz (IDL), Lisbon, Portugal. He is the head of SEGAL (Space & Earth Geodetic Analysis Laboratory), a collaborative project between UBI and IDL. He has been an active researcher in the use of GNSS for monitoring geophysical signals and for the definition of reference frames. In this respect, he has published several papers at peer-reviewed international journals and he is member of technical and scientific committees of EUREF and AFREF (European and African Reference Frames).

Chapter 6 "Earth Rotation" provides an overview of the state-of-the-art theoretical and observational aspects on Earth rotation. It is organised in five parts. The first section describes theoretical foundations of space-fixed and Earth-fixed reference systems, their mutual relation and the consequences of the implementation of the new IAU2000 resolutions. The second and third sections describe the results of astrometric and space geodetic observations of polar motion and length-of-day variations, respectively. The presented time-series are analysed in time and space with regard to signatures of gravitational and other geophysical processes in the Earth system. The fourth section deals with the physical foundations of Earth rotation models that are based on the balance of angular momentum in the Earth system. After theoretical considerations, various approaches for numerical Earth rotation models are presented. In Sect. 6.5, the chapter concludes with a discussion of the relation between modelled and observed variations of Earth rotation.

The author and co-author of Chap. 6 are Prof. Florian Seitz and Prof. Harald Schuh.

Florian Seitz studied geodesy at the Technische Universität München (TUM), Germany. After his graduation in the year 2000 he joined the Deutsches Geodätisches Forschungsinstitut (DGFI) in Munich, where he collaborated in various projects in the fields of Earth rotation, gravity field and surface geometry. In addition to theoretical studies, his main focus during his time at DGFI was the development of a numerical Earth system model for the simulation of atmospheric and hydrospheric effects on Earth rotation and gravity field, for which he obtained his doctorate from the TUM with distinction in 2004. During 2006 he joined NASA's Jet Propulsion Laboratory, Pasadena, USA, for a research visit for several months. He returned to the TUM as a professor for Earth Oriented Space Science and Technology in 2007. His main scientific interest is the integrated analysis of data of Earth observation satellites and space geodetic techniques and their application for numerical studies and models of the Earth system. At present he is chair of the study group SG-3 "Configuration Analysis of Earth Oriented Space Techniques" of IAG's Inter-commission Committee on Theory (2007–2011) and secretary of IAU Commission 19 'Rotation of the Earth' (2009–2012).

Harald Schuh is a full professor and Director of the Institute of Geodesy and Geophysics, Vienna University of Technology, Austria. Major areas of scientific interest are very long baseline interferometry (VLBI), Earth rotation, investigations of the troposphere and ionosphere. He graduated in 1979 from Bonn University, Germany and received his PhD in 1986. He occupied the following positions: Scientific assistant and associate professor at Bonn University (1980–1988); program scientist at the German Air and Space Agency (1989–1995); senior scientist and head of the Earth Rotation Division at DGFI, Munich (1995–2000); chair of the IVS Directing Board since 2007; president of IAU Commission 19 "Rotation of the Earth" (2009–2012); president of the Austrian Geodetic Commission since 2008 and president of the Austrian National Committee of the IUGG since 2009; member of the IAG executive committee and of various directing and governing boards; editorial board of the Journal of Geodesy (2003–2007); served as president, chair, member or consultant of various commissions, sub-commissions and working groups in geodesy (IAG) and astronomy (IAU); coordinator of the German Research Group on Earth Rotation (1999–2003); supervisor, co-supervisor, or examinator of more than 20 dissertations.

Chapter 7 is entitled "Equivalence of GPS Algorithms and its Inference". The equivalence principle of differential and un-differential GPS algorithms, combined and un-combined GPS algorithms as well as their mixtures are discussed. The principle can be alternatively argued as follows. As soon as the GPS data are measured, the information contents of the data are definitive ones. If the model used is the same and the principle of the adjustment and filtering is also the same, the obtained results should be equivalent. Advantages and disadvantages of different algorithms are relative and balanced. Based on the equivalence principle, the topic of independent parameterisation of the GPS observation model is discussed which points out where the singularity problem comes from. The consequences of the equivalence principle are important beyond the principle itself. The diagonalisation algorithm could be

extremely useful even for classic adjustment for reducing parameters. Separability of any observation equation and its normal equation may lead to an apparently unsolvable problem to be solvable or an accumulated one and later solvable one. Optimal criterion for ambiguity search may clear a decade-long confusion of the ambiguity searching criterion caused by the so-called LSSA method.

The author and co-authors of Chap. 7 are Dr. Guochang Xu (see Chap. 4), Prof. Yunzhong Shen, Prof. Yuanxi Yang (see Chap. 2), Prof. Heping Sun, Prof. Qin Zhang, Dr. Jianfeng Guo and Prof. Ta-Kang Yeh.

Yunzhong Shen is a professor in Department of Surveying and Geo-informatics Engineering of Tongji University. He received his Ph.D. from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences in 2001. He is now an editor of "Acta Geodetica et Cartographica Sinica". His main research interests are theory of geodetic data processing, satellite positioning and satellite gravimetry.

Heping Sun graduated in geophysics from University of Science and Technology of China in 1980. He obtained his doctorate from Catholique University of Louvain in Belgium in 1995. Having worked as a research assistant at the Institute of Seismology of the China Earthquake Prediction Administration in Wuhan from 1980 to 1991, Royal Observatory of Belgium from 1991 to 1996, he is a research professor in Institute of Geodesy and Geophysics, Chinese Academy of Sciences since 1997, and is director of the Institute since February 2005. He has been involved in gravity research, including theoretical study, data process and its application in Geodynamics; he has authored and co-authored more than 30 research papers.

Qin Zhang graduated in geodesy and survey engineering from Wuhan University in 1982 and 1994, respectively. She obtained her doctorate from Wuhan University in 2002. Having worked as a lecturer at the Wuhan University from 1982 to 1984 and as an associate professor at the Chang'an University, Xian, from 1984 to 2000, she works as a professor and vice dean at the Chang'an University since 2000. Prof. Zhang has been involved in GPS research since 1991 and has authored and co-authored several books. She is also an adjunct professor at Tianjin Institute of Urban Construction and an editor of some Chinese core journals. Several part-time positions are held by her, for example, as commissioner for Chinese Society for Geodesy, Photogrammetry and Cartography, executive commissioner and director for Society for Geodesy Photogrammetry and Cartography of Shaanxi province.

Jianfeng Guo is an associate professor at Information Engineering University (IEU), China. He obtained a B.Sc. in Mathematics from Xi'an Jiaotong University (XJTU) and an M.Sc. in Geodesy from IEU and a Ph.D. in Geodesy from Institute of Geodesy & Geophysics, Chinese Academy of Sciences (CAS). His research interests include geodesy and GNSS positioning and navigation.

Ta-Kang Yeh graduated in civil engineering and surveying engineering from National Chiao Tung University at Taiwan in 1997 and 1999, respectively. He also obtained his doctorate in geomatics from National Chiao Tung University at Taiwan in 2005. Having worked as an associate engineer at Industrial Technology Research Institute from 2000 to 2005, he has been an assistant professor at Ching Yun University since 2005. After working for 4 years he passed the promotion application and has been an associate professor since 2009. Moreover, he is the CEO of

e-GPS research center of Ching Yun University from 2008. He has been involved in GPS research since 1997 and has authored and co-authored several books and papers. He is also a member of International GNSS Service (IGS), International Association of Geodesy (IAG) and American Geophysical Union (AGU).

Chapter 8 "Marine Geodesy" presents an overview of geodetic contributions to the scope of the marine environment. After a brief introduction to the acquisition and use of hydrographic data the basic principles of hydroacoustics are presented. The importance of precise navigation is discussed and some examples are explained. The focus is put on the estimation of ship dynamic parameters and the contribution of geodesy to ship dynamics. A newly developed method for ship squat observation is described in detail which provides high precise data that allow discussing the correlation of trim and squat and furthermore the optimisation of ship underkeel clearance by considering the static trim and the squat-related dynamic trim change.

The author of Chap. 8 is Prof. Joerg Reinking. Joerg Reinking studied geodetic engineering at the Technical University (TU) Berlin, Germany, and received his diploma in 1988. Since 1988 he has worked as a research associate at TU Berlin and Technical University (TU) Braunschweig, Germany. He obtained his doctorate from TU Braunschweig in 1993 and worked as a scientist at the GeoForschungsZentrum (GFZ) Potsdam from 1993 to 1997. Since 1997 he has been a professor of geodesy, adjustment techniques and hydrographic surveying at the Jade University of Applied Sciences in Oldenburg, Germany. During the last decade he was engaged in the development of geodetic observation and analysis strategies for ship dynamic analysis (squat, trim and roll) and founded the Institute of Metrology and Adjustment Techniques and is a member of the Institute of Martime Studies in Elsfleth, Germany's largest nautical school.

Chapter 9 "Satellite Laser Ranging" introduces the reader to this space geodetic technique and covers the basics of instrumentation, error sources both in the measured and in calculated range, leading up to determination of observed-computed residuals, which provides an indication of "best-fit" orbit to the observations. Initially a range model is developed, which includes additional signal delays experienced by the transmitted laser pulse due to the atmosphere and general relativity. A description of centre-of-mass correction is given using LAGEOS as an example. Station range and time bias are discussed, highlighting the reasons for range bias variability while cautioning its application or interpretation as a station error without consideration of its diverse constituents. Following the measured range model, a simple orbit and force model are described, which includes the effects of gravity and its temporal changes, n-body perturbations, general relativity, atmospheric drag, solar and Earth radiation pressure as well as empirical forces. A calculated range model is then described, which makes allowance for station position variations due to solid Earth processes as well as other necessary adjustments. A brief overview of a typical SLR station is given, using MOBLAS-6 as an example. Operational aspects are covered with reference to the important role of the International

Laser Ranging Service (ILRS) and the global network of participating SLR stations.

The author of Chap. 9 is Dr. Ludwig Combrinck. Ludwig Combrinck is employed at the Hartebeesthoek Radio Astronomy Observatory (HartRAO) located near Krugersdorp, South Africa. HartRAO is a facility of the National Research Foundation (NRF). Ludwig was awarded a PhD by the University of Cape Town in 2000, his thesis focussed on GNSS applications for precise positioning. He is responsible for the Space Geodesy Programme at HartRAO, which includes the NASA satellite laser ranging station, MOBLAS-6. In 2009 he was appointed professor-extraordinaire at the University of Pretoria and research associate at the University of South Africa where he lectures part-time. His main research interests currently include applications of space geodetic techniques, reference frame development for Africa and the development of a new high-accuracy satellite and lunar laser ranger for South Africa. His diverse interests in the applications of space geodesy have resulted in the establishment of geodetic stations throughout Africa, Marion Island and Antarctica, in collaboration with international partners.

Chapter 10 "Superconducting Gravimetry" is related to measuring, evaluation and interpretation of superconducting gravimeter data. It gives an overview of the instrument, the data processing techniques including pre-processing and Earth tide analysis and its application in geodynamics, combined with the correction of environmental influences (atmosphere, hydrosphere and ocean). The corresponding sections of this chapter include the description of the instrument, site selection and observatory design, calibration of the gravity sensor, noise characteristics, description and modelling of the principal constituents of the gravity signal, analysis of different surface gravity effects, combination of ground-and satellite-derived gravity variations, co-seismic gravity changes, up to future applications.

The author of Chap. 10 is Dr. Jürgen Neumeyer. Jürgen Neumeyer graduated in electrical engineering at Technical University Ilmenau in 1965. He obtained his first Ph.D. in electrical engineering at University of Ilmenau in 1971 and his second Ph.D. in geophysical measurement technique at Academy of Sciences of GDR in 1989. Since 1978 he has been dealing with geo-sciences. He worked from 1978 to 1991 at "Central Institute Physics of the Earth" Potsdam in the fields of gravimetry, seismology and remote sensing. From 1992 to 2007 he was working at "GeoForschungZentrum Potsdam" in the field of superconducting- and airborne-gravimetry and GPS. During this time he published his scientific results in several papers.

Chapter 11 "Synthetic Aperture Radar Interferometry" introduces the principles and data processing of the SAR interferometry including differential SAR interferometry, corner reflector SAR interferometry (CR-INSAR) and some of the practical applications. In Sect. 11.2 the basics of the SAR imaging are briefly reviewed to understand the SAR imaging process and SAR image feature, which also is the background of the SAR interferometry. Section 11.3 describes the principle and data processing of the SAR interferometry for digital elevation model (DEM)

generation. Section 11.4 deals with the differential SAR interferometry. In Sect. 11.5 the differential interferometry of the persistent coherent is discussed.

The author of Chapter 11 is Dr. Ye Xia. Ye Xia received the Dr.-Ing. degree in navigation from the University of Stuttgart, Germany, in 1995, the M.S. degree in electrical engineering from Hunan University, China, in 1982, and the B.A. degree in electrical engineering from Shanghai Jiao Tong University, China, in 1968. He is currently a senior scientist at the Geo-Research Center Potsdam, Germany. His research interests include electrocircuit theory, active filter design, imaging and interferometry of the synthetic aperture radar and the INSAR applications in geography survey and geological disasters monitoring.

The book has been subjected to an individual review of chapters. I am grateful to reviewers Prof. Aleksander Brzezinski of the Space Research Centre of the Polish Academy of Sciences, Prof. Wu Chen of HongKong Polytech University, Prof. Alexander Härting of the University of Applied Sciences Oldenburg, Prof. Urs Hugentobler of Technical University Munich, Dr. Corinna Kroner, Dr. Svetozar Petrovic and Dr. Ludwig Grunwaldt of GFZ, Prof. Xiaohui Li of National Time Service Center in Xi'an, Prof. Zhiping Lü and Dr. Jianfeng Guo of Information Engineering University (IEU) in Zhengzhou, Prof. Yunzhong Shen of Tonji University in Shanghai, Prof. Heping Sun and Prof. Jikun Ou of the Institute of Geodesy and Geophysics (IGG) in Wuhan, Dr. Tianhe Xu of GFZ and the Institute of Surveying and Mapping (ISM) in Xi'an, Prof. Ta-Kang Yeh of Ching Yun University of Taiwan, Dr. Walter Zürn of University Karlsruhe. As editor of this book I made a general review of the whole book. A grammatical check of technical English writing has been performed by Springer Heidelberg.

I wish to thank sincerely the key authors of the individual chapters: Dr. Ludger Timmen of University Hannover, Prof. Yuanxi Yang of ISM in Xi'an, state scientist Rene Forsberg and Dr. Arne V. Olesen of Danish Space Center in Copenhagen University, Dr. Luisa Bastos and Dr. Machiel Bos of University of Porto, Dr. Rui Manuel Fernandes of University of Beira Interior (UBI), Prof. Florian Seitz of Technische Universität München, Prof. Harald Schuh of Vienna University of Technology, Prof. Yunzhong Shen of Tonji University in Shanghai, Prof. Heping Sun of IGG in Wuhan, Prof. Qin Zhang of ChangAn University in Xi'an, Dr. Jianfeng Guo of IEU in Zhengzhou, Prof. Ta-Kang Yeh of Ching Yun University of Taiwan, Prof. Joerg Reinking of University of Applied Sciences in Oldenburg, Dr. Ludwig Combrinck of Hartebeesthoek Radio Astronomy Observatory, Dr. Jürgen Neumeyer of Potsdam, Dr. Ye Xia of GFZ. Without their consistent efforts such a book will be never available. I also wish to thank sincerely scientists who made great efforts for enriching this book. They are Prof. Jürgen Kusche of University Bonn, Dr. Oscar Colombo of NASA and Prof. Tianyuan Shih of Central University of Taiwan.

I wish to thank sincerely the former directors Prof. Dr. Ch. Reigber and Prof. Dr. Markus Rothacher of GFZ for their support and trust during my research activities at the GFZ and for granting me special freedom of research. I also wish to thank sincerely Prof. Yuanxi Yang of ISM in Xi'an, Prof. Qin Zhang of ChangAn

University in Xi'an, Prof. Heping Sun, Prof. Jikun Ou and Prof. Yunbin Yuan of IGG in Wuhan for their friendly support by organising the International Geodetic Forum Xi'an 2006, which is the origin of the idea to write and edit such a scientific book. The Chinese Academy of Sciences is thanked for the Outstanding Overseas Chinese Scholars Fund, which supported greatly the valuable scientific activities.

Special thanks goes to Springer, Heidelberg; their support and their evaluation for such a book are preconditions for successfully organizing this publication. I am also grateful to Dr. Chris Bendall of Springer, Heidelberg, for his valuable advice.

Potsdam, Germany Guochang Xu
June 2009

Contents

Contributors

Luisa Bastos Observatório Astronómico, Faculdade de Ciências, Universidade do Porto, Alameda do Monte da Virgem, 4420-146 V. N. Gaia, Portugal, lcbastos@fc.up.pt

Machiel Bos Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal, mbos@ciimar.up.pt

Ludwig Combrinck Space Geodesy Programme, Hartebeesthoek Radio Astronomy Observatory, Krugersdorp 1740, South Africa, ludwig@hartrao.ac.za

Rui Manuel Fernandes Universidade da Beira Interior, Covilhã, Portugal and Instituto Geofísico Infante D. Luíz Lisboa, Portugal, rmanuel@di.ubi.pt

Rene Forsberg National Space Institute, Technical University of Denmark, Juliane Maries Vej 30, Copenhagen 2100, Denmark, rf@space.dtu.dk

Jianfeng Guo Information Engineering University of Zhengzhou, Zhengzhou, P.R. China, jianfeng.guo@gmail.com

Jürgen Neumeyer GeoForschungZentrum Potsdam, Potsdam, Germany Juergen_Neumy@yahoo.de

Arne V. Olesen National Space Institute, Technical University of Denmark, Juliane Maries Vej 30, Copenhagen 2100, Denmark avo@space.dtu.dk

Joerg Reinking Jade University of Applied Sciences, Wilhelmshaven/Oldenburg/Elsfleth, Germany, reinking@jade-hs.de

Harald Schuh Institute of Geodesy and Geophysics, Vienna University of Technology, Wien, Austria, harald.schuh@tuwien.ac.at

Florian Seitz Earth Oriented Space Science and Technology, Technische Universität München (TUM), Arcisstr. 21, D-80333 Munich, Munich, Germany, seitz@bv.tum.de

Yunzhong Shen Department of Surveying and Geomatics, TongJi University, Shanghai 200092, P.R. China, yzshen@mail.tongji.edu.cn

Heping Sun Institute of Geodesy and Geophysics, CAS, Wuhan 430077, P.R. China, heping@asch.whigg.ac.cn

Ludger Timmen Institut für Erdmessung (IfE), Leibniz Universität Hannover (LUH), Schneiderberg 50, Hannover 30167, Germany, timmen@ife.uni-hannover.de

Ye Xia German Research Centre for Geosciences, Telegrafenberg A17, Potsdam D-14473, Germany, xia@gfz-potsdam.de

Guochang Xu GFZ German Research Centre for Geosciences, Department 1: Geodesy and Remote Sensing, Telegrafenberg, 14473 Potsdam, Germany, xu@gfz-potsdam.de

Yuanxi Yang Xian Research Institute of Surveying and Mapping, Xian 710054, China, yuanxi@pub.xaonline.com

Ta-Kang Yeh Institute of Geomatics and Disaster Prevention Technology, Ching Yun University, Jhongli 320, Taiwan, bigsteel@cyu.edu.tw

Qin Zhang Department of Geology and Surveying, ChangAn University, XiAn 710054, P.R. China, zhangqinle@263.net.cn

Chapter 1
Absolute and Relative Gravimetry

Ludger Timmen

Contents

L. Timmen (✉)
Institut für Erdmessung (IfE), Leibniz Universität Hannover (LUH), Schneiderberg 50, 30167 Hannover, Germany
e-mail: timmen@ife.uni-hannover.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_1,

1.1 Introduction

Absolute and relative gravimetry allow the determination of gravity acceleration, usually just called gravity, for specific positions as well as the detection of gravity changes with time at a given location. For high-accuracy demands, the geometrical position of a gravity point has to be defined very accurately, e.g. in geodynamic research projects, at a height along the vertical above a ground mark. Geodetic networks with local, regional or global extent can be surveyed to monitor short-term and long-term gravity variations.

This chapter refers particularly to experience gained at the Institut für Erdmessung (IfE), Leibniz Universität Hannover (LUH). In the following, an overview of relative and absolute gravimetry (instrumental techniques, observation equations, accuracies, etc.) is given. Exemplarily for present state-of-the-art absolute and relative gravimeters, the main characteristics and accuracy estimates for the Hannover instruments are presented.

Because of the dynamics within the Earth's system (tectonics, climate change, sea-level rise), the national and international base networks are not stable with time. With the high accuracies of modern geodetic techniques, combined with the high quality of the base net stations (stable environment, customised facilities), the networks serve more and more as control systems for environmental changes and surface deformations.

The recommended unit of acceleration in the Système International d'Unités (SI) is the unit $\mathrm{m/s^2}$ (BIPM 2006). In geodesy and geophysics, the non-SI unit Gal ($1\ \mathrm{Gal} = 1\ \mathrm{cm/s^2} = 0.01\ \mathrm{m/s^2}$) is also used to express acceleration due to gravity. In order to provide gravity differences and to describe small deviations or uncertainties of the measurements, the following units are helpful:

$$1\ \mathrm{mGal} = 10^{-5}\mathrm{m/s^2} = 10^{-6}g \quad \text{and} \quad 1\ \mu\mathrm{Gal} = 10^{-8}\mathrm{m/s^2} = 10^{-9}g\,. \quad (1)$$

1.2 Characteristics of Absolute Gravimetry

1.2.1 General Aspects

To realise the advantages of absolute gravimetric measurements, some particular features of the gravity acceleration g, usually just called gravity g, for a defined geometrical point should be explained first. The gravity acceleration at a surface point depends on the following:

1. The position relative to the Earth's masses and their density distribution (integral effect caused by the gravitational force of the Earth's masses)
2. The position relative to the Earth's rotation axis (effect caused by the centrifugal force due to the Earth's rotation)

The g-value of a point at the Earth's surface (e.g. bench mark attached to a pier) changes with the following:

- Varying distance to the centre of masses of the Earth (geocentre) caused by vertical movements of the measuring point, e.g. due to crustal deformations, and by secular variations of the position of the geocentre (subtle effect, requires long-time measuring series)
- Mass shifts and redistributions within the system Earth (including atmosphere and hydrosphere) and especially with near-surface variations within the crust (e.g. groundwater changes, sediment compaction)
- Changing distance to the Earth's rotation pole due to lateral movements (subtle effect, e.g. due to plate tectonics)

Absolute gravity measurements are most sensitive to height changes and provide an obvious way to define and control the vertical height datum. No additional reference points (connection points) at the Earth's surface and no observations of celestial bodies (quasars, stars, planets, moon) or satellites are needed. Shortcomings of relative gravimetry, like calibration problems and deficiencies in the datum level definition, can be overcome. The accuracy of an absolute gravity net is independent of geographical extension which allows applications on local, regional and global scales with consistent measurement quality. An independent verification of displacements measured geometrically with GPS (Global Positioning System), VLBI (Very Long Baseline Interferometry) and SLR (Satellite Laser Ranging) is possible. A combination of gravimetric and geometric measurements may enable discrimination among subsurface mass movements associated with or without a surface deformation.

1.2.2 Objectives of Geo-scientific and State-geodetic Surveys

The benefit of absolute gravimetry has already been exploited in different scientific projects. The International Absolute Gravity Basestation Network (IAGBN) serves, among other purposes, for the determination of large-scale tectonic plate movements (Boedecker and Fritzer 1986; Boedecker and Flury 1995). The recommendations of the Interunion Commission on the Lithosphere on mean sea level and tides propose the regular implementation of absolute gravity measurements at coastal points, 1–10 km away from tide gauges (Carter et al. 1989). The height differences between gravity points and tide gauges have to be controlled by levelling or GPS. In Great Britain, the main tide gauges are controlled by repeated absolute gravity determinations in combination with episodic or continuous GPS measurements (Williams et al. 2001). Torge (1998a, b) describes the changing role of gravity reference networks due to the modern approach to realising the network standards by absolute observations.

Overall, absolute gravimetry can be an important research tool for studying geodynamic processes, especially land uplift effects due to postglacial rebound

(PGR). Lambert et al. (1996) give an overview of the capability of absolute gravity measurements in determining the temporal variations in the Earth's gravity field. In Lambert et al. (2001), the gravimetric results for the research of the Laurentide post-glacial rebound in Canada are described. Mäkinen et al. (2007) compare observed gravity changes in Antarctica with modelled predictions of the glacial isostatic adjustment as well as of the glacier mass balance.

Since 1986, several gravimetric projects were performed by IfE with the absolute gravimeters JILAg-3 (e.g. Torge 1990, 1993; Timmen 1996) and FG5-220 (Gitlein et al. 2008; Timmen et al. 2006a). These activities served the following main objectives:

- Establishing and improving international and national gravity reference networks to realise a homogeneous gravity standard (datum definition in level and scale) of regional to global extent; calibration systems for relative gravimetry are needed
- Installing and strengthening regional and local networks in tectonically active areas with absolute gravimetric measurements and following re-observations; such monitoring systems may serve for geophysical research on the mechanism of crust formation and on the rheology of Earth's mantle and crust
- Monitoring the vertical stability of tide gauge stations to separate sea-level changes from land surface shifts; this serves to constrain parameters related to global climatic change

With the initiation of the GRACE satellite experiment (Gravity Recovery and Climate Experiment, e.g. Wahr and Velicogna 2003; Tapley et al. 2004), a new requirement has arisen for absolute gravimetry:

- To provide the most accurate "ground truth" for GRACE

The results from both data sets describe changes of the gravity field at the Earth's surface or at the geoid. The terrestrial data can not only be used to validate the GRACE products (Müller et al. 2006) but may also serve as a completion of the satellite results.

In the future, two additional tasks may become important applications:

- Monitoring of human-caused changes in aquifers and deep water reservoirs by water extraction
- Contributing to the definition of ground-based geodetic reference networks within the activities for the Global Geodetic Observing Systems (GGOS) of the International Association of Geodesy (IAG)

GGOS will provide the observational basis to integrate the different geodetic techniques. The purpose of the globally collected geodetic data is to collate and

analyse information about global processes and changes which are important for world societies. An overview and further details about GGOS can be obtained from Pearlman et al. (2006). In Ilk et al. (2005), detailed information about mass transport processes in the Earth system is given.

1.3 Measurements with Free-Fall Absolute Gravimeters

In January 1986, the Institut für Erdmessung (IfE), Leibniz Universität Hannover (LUH), received the absolute gravimeter JILAg-3 which was the first transportable system located in Germany (Torge et al. 1987). The free-fall system was developed at the Joint Institute of Laboratory Astrophysics (JILA, Faller et al. 1983) of the University of Colorado. The so-called JILAg-3 was the third gravimeter of a series of six JILA instruments and was successfully employed by IfE in more than 130 absolute gravity determinations at about 80 different stations (South America, China, Greenland, Iceland, central and northern Europe). In December 2002, IfE had received a new FG5 absolute gravity meter (FG5-220) from Micro-g Solutions, Inc. (now Micro-g LaCoste, Inc., Erie, Colorado), which was a state-of-the-art instrument (Niebauer et al. 1995) and replaced the older JILAg-3. Based on the JILA design, the FG5 generation has overcome several constructively predefined shortcomings and represents an essential improvement in operation and accuracy. The first fully operational FG5 instrument (FG5-101) was already available in 1993, manufactured by AXIS Instruments Company in Boulder, CO (Carter et al. 1994). The achievement of AXIS became possible after the National Institute of Standards and Technology (NIST, Boulder, USA) and the former Institute of Applied Geodesy (IfAG, now Federal Agency for Cartography and Geodesy, BKG, Frankfurt, Germany) joined forces in 1990 to produce an advanced instrument capable of providing more stringent data constraints on geophysical investigations.

The FG5 series represents the currently most advanced instruments and has to be assumed as the best instrumental realisation to measure the absolute gravity acceleration. Besides the FG5 meter for most accurate applications, a portable absolute gravimeter for harsh field environments has been developed by Micro-g LaCoste, Inc., called A10 (Niebauer et al. 1999, see also the Micro-g LaCoste internet pages). This unique instrument allows a data sampling rate of 1 Hz and provides a precision of 10 μGal after 10 min of measurements at a quiet outdoor field site. An absolute accuracy of 10 μGal can be achieved for a station determination. Liard and Gagnon (2002) tested their new A10 in 2001 at the International Comparison of Absolute Gravimeters in Sèvres, France. The investigations of Schmerge and Francis (2006) confirm the accuracy specifications of the manufacturer.

Figure 1.1 shows the two types of the Hannover absolute gravimeters, the instruments JILAg-3 and FG5-220. During the period from 1986 to 2000, the JILAg-3 gravimeter was used by IfE for absolute gravity determinations on more than 80 different sites worldwide. The measurements with the presently employed FG5-220

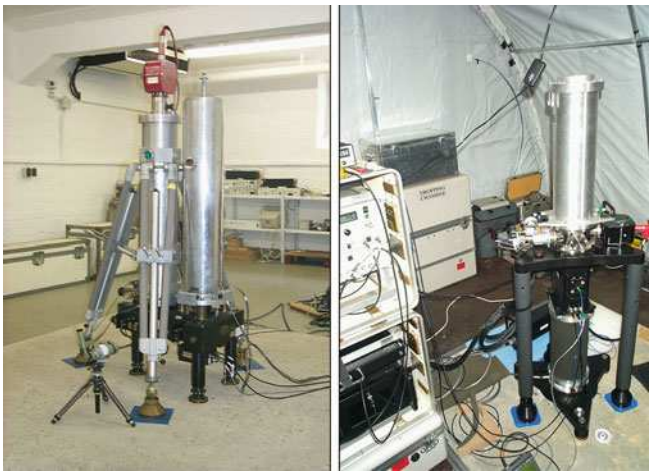


Fig. 1.1 The two absolute gravimeters of the Leibniz Universität Hannover: *left* JILAg-3 employed from 1986 to 2000 (here reference measurements in Hannover), *right* FG5-220 operated since 2003 (tent measurements in Denmark)

started in 2003, and more than 40 different sites in central and northern Europe have already been visited.

1.3.1 Principles of FG5 Gravimeters

Modern absolute gravity measurements are based on time and distance measurements along the vertical to derive the gravity acceleration at a specific position on the Earth, cf. Torge (1989). The expression "absolute" is based on the fact that the time and length standards (rubidium clock, helium–neon laser) are incorporated as components of the gravimeter system. No external reference like a connecting point is required. The FG5 series is presently the most common gravimeter model, which may be considered as the successor system of the JILA generation (Carter et al. 1994; Niebauer et al. 1995). The influence of floor vibration and tilt on the optical path could largely be removed by the improved interferometer design. The iodine-stabilised laser, serving as the primary length standard, is separated from the instrumental vibrations, caused by the free-fall experiments (drops), by routing the laser light through a fibre optic cable to the interferometer base; see Fig. 1.2.

During a drop, the trajectory of a test mass (optical retro-reflector) is traced by laser interferometry over the falling distance of about 20 cm within an evacuated chamber. The "co-falling" drag-free cart provides a molecular shield for the dropped object. The multiple time–distance data pairs collected during the drop (FG5: 700 pairs at equally spaced measuring positions, JILAg: 200) are adjusted to a fitting curve (almost parabolic) as shown in Fig. 1.3, giving the gravity acceleration g for the reference height above floor level (FG5: ~1.2 m, JILAg: ~0.8 m).

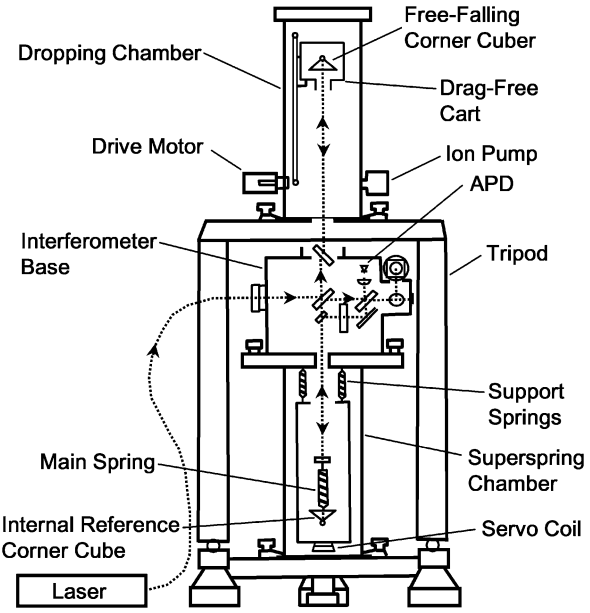


Fig. 1.2 Schematic diagram of the FG5 absolute gravimeter, after Micro-g Solutions, Inc. (1999), courtesy of Micro-g Lacoste, Inc.

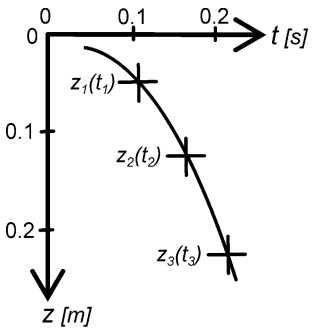


Fig. 1.3 The free-fall in a time–distance diagram

1.3.2 Observation Equation

In a uniform gravity field, the motion of a freely falling mass m can be expressed with the following equation of motion:

$$m\,\frac{d^2z}{dt^2} = m\,\ddot{z} = m\,g. \tag{2}$$

Figure 1.3 shows the time–distance diagram with the axis t and z where the z-axis coincides with the direction of gravity. Eliminating m in (2), the integration yields an equation for the velocity

$$\dot{z} = \dot{z}_0 + g\,t, \quad \text{with } \dot{z}_0 = v_0, \tag{3}$$

and thereafter for the position

$$z = z_0 + v_0\,t + \frac{g}{2}t^2 \tag{4}$$

of a body during a free fall. The initial parameters, displacement z_0 and velocity v_0, are adjustment unknowns valid at the time $t = 0$. For most accurate gravity determination, the non-uniformity of the Earth's gravitational field has to be taken into account. Along the plumb line, the gravity acceleration g varies with height. This can be considered as a linear gravity change along the free-fall trajectory during an experiment with an absolute gravimeter. Hence formula (2) is extended with a constant vertical gravity gradient γ:

$$m\ddot{z} = m(g_0 + \frac{dg}{dz}z) = m(g_0 + \gamma z). \tag{5}$$

The acceleration g_0 is defined for the position $z = 0$ which is, in common practice with FG5 and JILA meters, the resting position of the gravity centre of the test mass at the start of the free-fall experiment ("top-of-the-drop"). Neglecting the initial parameters and double integration of (5) gives (Cook 1965)

$$z = \frac{g_0}{\gamma}(\cosh \gamma^{1/2} t - 1), \quad \text{with} \quad z_0 = v_0 = 0\,. \tag{6}$$

Because the initial parameters z_0 and v_0 have to be included, the variable z is expanded to the power series

$$z = f(t) = c_0 + c_1(t - t_0) + c_2(t - t_0)^2 + c_3(t - t_0)^3 + \cdots\,. \tag{7}$$

With $t_0 = 0$ the following equations are deduced:

$$z = c_0 + c_1\,t + \ c_2\,t^2 + \ c_3\,t^3 + \cdots, \quad z(t = 0) = c_0 \quad \Rightarrow\ c_0 = z_0, \tag{8}$$

$$\dot{z} = \quad c_1 \ + 2c_2 t \ + 3c_3 t^2 + \cdots, \quad \dot{z}(t = 0) = c_1 \quad \Rightarrow\ c_1 = v_0, \tag{9}$$

$$\ddot{z} = \quad 2c_2 \ + 6c_3 t \ + \cdots\,. \tag{10}$$

Inserting these series in (5) yields

$$2c_2 + 6c_3 t + 12c_4 t^2 + 20c_5 t^3 + \cdots = g_0 + \gamma(c_0 + c_1 t + c_2 t^2 + c_3 t^3 + \cdots)\,. \tag{11}$$

Comparing the coefficients on the left side of the equation with the right side, the constants are obtained as

$$c_2 = \frac{1}{2}g_0 + \frac{1}{2}z_0\gamma\ , \quad c_3 = \frac{1}{6}\gamma\, v_0\ , \quad c_4 = \frac{1}{24}\gamma\, g_0 + \frac{1}{24}\gamma^2 z_0\ , \quad \ldots . \tag{12}$$

Considering the terms up to the order t^4, (8) can be re-written as

$$z(t) = z_0\left(1 + \frac{1}{2}\gamma\, t^2 + \frac{1}{24}\gamma\, t^4\right) + v_0\left(t + \frac{1}{6}\gamma\, t^3\right) + \frac{1}{2}g_0\left(t^2 + \frac{1}{12}\gamma\, t^4\right)\ . \tag{13}$$

Equation (13) is the observation equation which is used in absolute gravimetry to derive a g-value from the multiple time–distance measurement pairs in a least-squares adjustment. Because of its subtle contribution, the t^4 term in the z_0-dependent expression can be neglected. The finite velocity of light c must be taken into account by adding the term z/c to the observed (raw) time values t' before the least-squares adjustment is carried out:

$$t = t' + \frac{z}{c}. \tag{14}$$

The reference height (position $z = 0$) of the derived free-fall acceleration g_0 depends on the setup of the instrument and should be defined by the operators with an accuracy of ± 1 mm to preserve the accuracy of the measurement system. For further theoretical considerations about the equation of motion in absolute gravimetry, it is recommended to study, e.g. Cook (1965) and Nagornyi (1995).

In Torge (1993), a simple formula is given to assess the required accuracy for the time and distance measurements:

$$\frac{dg}{g} = \frac{dz}{z} - 2\frac{dt}{t}. \tag{15}$$

Equation (15) is obtained by the differentiation of (4) and setting z_0 and v_0 to zero. Asking for a relative accuracy $dg/g = 10^{-9}$ ($\equiv dg = 1\,\mu$Gal) and considering a falling path of 0.2 m with a falling time of about 0.2 s, the accuracy requirement for the time and distance measurements is 0.2 nm and 0.1 ns, respectively. For state-of-the-art absolute gravimeters, this accuracy level is provided by the simultaneously performed atomic time and laser interferometric distance measurements.

1.3.3 Operational Procedures with FG5-220

Within the operational procedures with FG5-220, as employed at IfE, the time interval between two drops is 10 s which includes the reset of the falling corner cube and the online adjustment. For the reduction of local noise and other disturbances, 1,500–3,000 computer-controlled drops are performed per station determination. Generally, the measurements are subdivided into sets of 50 drops each and distributed over 1–3 days. The result of a station determination is the average of all

drops, reduced for gravity changes due to Earth's body and ocean tides, polar motion and atmospheric mass movements, as explained in Sect. 1.5.

Relative gravimetric measurements are still highly important to transfer the absolute gravimetry results to network points at floor level or to another height level along the vertical that has been agreed on, e.g. for comparisons of different absolute gravity determinations. However, to preserve the accuracy of the absolute measurements for present and future investigations and applications, the absolute gravity result should not be affected by uncertainties in the vertical gradient due to measurement errors from relative gravimetry or deteriorated by unknown non-linearities in the gradient (Timmen 2003). These demands are fulfilled by defining the reference height close to a position where the influence of an uncertainty in the vertical gravity gradient becomes almost zero ("dead-gradient-point"); see Fig. 1.4. The corresponding position is approximately one-third of the falling distance below the first measured position of the free-fall trajectory as used in the adjustment computation (FG5-220: ~1.21 m above floor level). Therefore, all gravity determinations with the current Hannover FG5 instrument are referred not only to the ground floor mark but also to the reference height of 1.200 m above floor level or above the ground mark.

For the reduction of the absolute gravity value to the floor mark, the observed gravity difference (hereafter called gradient) is needed. Following the IfE standard procedure, the vertical gravity gradient is determined with two LaCoste and Romberg gravimeters with integrated feedback systems (Schnüll et al. 1984) or with a Scintrex Autograv CG3M (since 2002) using a tripod of about 1 m height. By observing the difference 10 times with each relative meter, the gravity

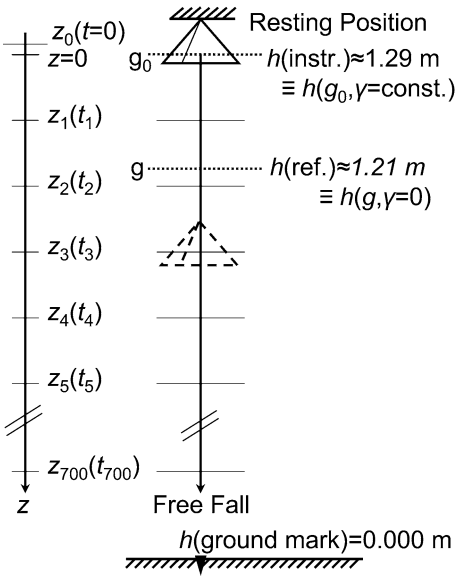


Fig. 1.4 Depending on the setup, the instrumental height of FG5-220 is around 1.29 m (top-of-the-drop). For geodynamic research, the g-value is referred to the fixed height 1.200 m above measuring mark to avoid uncertainties from the gradient assumption

difference is normally obtained with a standard deviation of about 1 μGal. Referring the gravity difference to a height difference of 1.000 m, the vertical gravity gradient γ is obtained. Here, a linear gravity change with height is assumed (constant gradient). For geodynamic research, often a more accurate knowledge about the vertical gravity change is required by considering a second-degree polynomial for the height dependency. In those cases, gravity differences Δg are measured between variable height levels h above the ground mark (cf. Fig. 1.5). A least-squares adjustment of observation equations provides an overconstrained solution for the coefficients γ_1 and γ_2 describing the linear and quadratic part of the polynomial:

$$\Delta g(h_i, h_j) = \gamma_1(h_j - h_i) + \gamma_2(h_j^2 - h_i^2). \tag{16}$$

With (16), an observed absolute gravity value with its defined reference height can be referred to any position within the perpendicular above the ground mark up to about 1.5 m (highest relative gravity measurement position).

For the site selection, preferences are given to buildings with a stable environment inside the observation room (stable temperature, no direct sun, relative humidity below 70%) and a solid foundation like a concrete pier, a reinforced concrete base slab or a concrete slab attached to bedrock.

Fig. 1.5 Measurement of the non-linear vertical gravity change along the perpendicular with a Scintrex relative gravimeter. Tripods are used for variable setup heights

1.3.4 Accuracy and Instrumental Offset

The manufacturers of the JILAg and of the FG5 systems performed an error budget analysis to determine the single instrumental uncertainty contributions through calculations and measurements of known physical effects. Niebauer (1987) derived a total error of 3 μGal for JILAg instruments. In Niebauer et al. (1995) a total uncertainty of 1.1 μGal is obtained from the FG5 instrumental error budget (Table 1.1).

To assess the accuracy of the transportable absolute gravimeters from the user point of view, the experiences with the Hannover instruments JILAg-3 and FG5-220 are used to derive an empirical accuracy estimate. For both instruments, the accuracy and stability have been continuously controlled by comparisons with other absolute gravity meters and with repeated measurements in several stations after time intervals of some months to a few years. A rigorous control of the absolute accuracy with respect to a "true" gravity value at the moment of an absolute gravity measurement is not possible. The real g-value with a superior accuracy is not known, and a "standard" absolute gravimeter which is superior to the state-of-the-art FG5 meters does not exist. Therefore, the empirical accuracy estimates have to

Table 1.1 Instrumental error budget of FG5 gravimeter and gravitational "noise" due to incomplete modelling and reduction of gravity variations with time, after Niebauer et al. (1995)

FG5 error source	Uncertainty [μGal]
Residual air pressure	0.1
Different temperature	0.1
Magnetic field gradient	0.1
Electrostatics	0.1
Attraction of apparatus	0.1
Verticality	0.1
Air-gap modulation	0.6
Laser wavelength	0.1
Corner-cube rotation	0.3
Coriolis effect	0.4
Floor recoil and tilt	0.1
Electronic phase shift	0.6
Frequency standard	0.2
Glass wedges	0.3
Diffraction limit	0.2
Total uncertainty (r.m.s.)	1.1
Gravitational "noise"	
Solid Earth tides	0.2–0.5
Equilibrium ocean load	0.2
Tidal swell and surge	5
Atmospheric attraction and loading	1–5
Groundwater table variations	Site-dependent
Polar motion	0.01
Microseisms (<100 Hz)	0

be understood as describing the agreement of the instruments' measuring level and their time stability with regard to the international absolute gravity datum definition. Here, the international datum is defined by the physical standards (time and length) and, in addition, as the average result obtained from all operational absolute gravimeters participating in the international comparison campaigns.

For JILAg-3, Torge (1991) estimated the short-term and long-term accuracy of a station determination between 5 and 10 μGal. On average, an accuracy estimate of 7 μGal was obtained. The instrumental precision by itself is assumed to be 4–5 μGal, which does not consider errors introduced by real gravity changes, e.g. due to subsurface water variation. For FG5-220, a realistic mean accuracy estimate seems to be about 3 μGal (cf. Timmen et al. 2006b; Francis and van Dam 2006; Francis et al. 2010; Bilker-Koivula et al. 2008). These empirical estimates incorporate

- Instrumental errors, e.g. due to instrumental vibrations or laser instabilities
- Gravitational "noise" due to incomplete modelling and reduction of gravity variations with time

Because most of the IfE measurements serve for local and regional gravimetric control, especially for geodynamic investigations in tectonically active areas, the long-term measuring stability of the two gravimeters is a major concern. To compare the results of JILAg-3 with recent observations of FG5-220, no systematic difference due to the gravimeters themselves should exist, or the instrumental offset should be well known. Within this context the instrumental offset should be understood as a mean measuring offset (bias) valid for a long time period, e.g. some years or even the gravimeters' lifetime. One possibility for detecting such an offset is to compare observation series of both instruments performed at a reference station where long-term stable gravity acceleration can be assumed (no significant secular change). The JILAg-3 reference station Clausthal in the Harz Mountains (stable bedrock) was occupied by FG5-220 at four different epochs in 2003 (January, May, June and October) to derive a reliable mean g-value for 2003 which is only slightly affected by seasonal hydrological changes. In Table 1.2, the mean result is compared with the mean from 29 gravity determinations with JILAg-3 performed in the

Table 1.2 Mean gravity values for station Clausthal (Germany) derived with JILAg-3 ($n = 29$ occupations, 1986–2000) and FG5-220 ($n = 4$ in 2003). The given s_i are standard deviations for a single gravity determination

JILAg-3/FG5-220 comparison	Remarks	Gravimeter	Period	Mean g-result [μGal]
Clausthal (Harz mountains)	IfE reference station for JILAg-3, reference height 0.000 m	JILAg-3	1986–2000	981,115,734.5 $s_i = 4.9$, $n = 29$
		FG5-220	January–October 2003	981,115,725.1 $s_i = 2.3$, $n = 4$ $\Delta g = +9.4$

Table 1.3 Gravity differences (JILAg-3 minus FG5-101) obtained from the International Comparisons of Absolute Gravimeters (ICAG) in Sèvres 1994 and 1997 and during the surveying of the national German base network DSGN94 (five identical stations) and from three comparisons at the Clausthal reference station, after Torge et al. (1999)

Comparisons of JILAg-3 (IfE) and FG5-101 (BKG)	Discrepancy [μGal]
ICAG94, BIPM, pier A0	+9.0
ICAG97, BIPM, pier A	+8.1
DSGN94	+8.2
Clausthal reference station	+9.4

period 1986–2000. The standard deviation of the mean values is in both cases about 1 μGal. An obtained discrepancy of +9.4 μGal indicates a significant offset between the measuring levels of these two absolute gravimeters. Similar discrepancies have also been reported by Torge et al. (1999) when comparing measurements from FG5-101 (BKG) and JILAg-3 performed in the years 1994–1997. These comparisons showed a discrepancy varying between +8.1 and +9.4 μGal (Table 1.3). Figure 1.6 shows the time series of absolute gravity determinations at station Clausthal (point 522) observed with the two Hannover instruments (offset correction applied). The decline in the four observed g-values at the Clausthal station in 2003 should be connected to the very dry season in northern Germany. A similar but much stronger gravity change was measured in Hannover when the groundwater table fell 70 cm accompanied by a gravity decrease of about 13 μGal, see also Sect. 1.6.1.

By taking the offset correction of −9 μGal into account for all JILAg-3 observations, a stable measurement level for a time span of more than 20 years is assumed to be available with the two Hannover instruments. This is in accordance with the present knowledge that the FG5 series is presently the best instrumental realisation of absolute gravimeters. Nevertheless, to meet the accuracy requirements for long-term research over many decades and for comparability with other instruments, the observation level of the JILAg-3–FG5-220 couple has to be verified by comparisons with other absolute gravimeters. Since the 1980s,

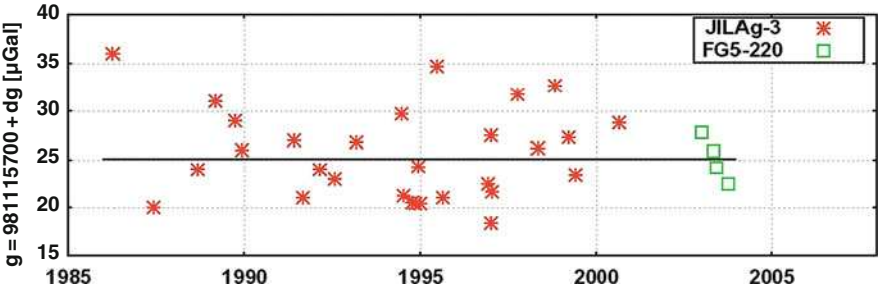


Fig. 1.6 Absolute gravity determinations with JILAg-3 and FG5-220 at station Clausthal (CLA522, trend −0.1 ± 0.2 μGal/year). An instrumental offset of −9 μGal (±1 μGal) was applied to the JILAg-3 results

International Comparisons of Absolute Gravimeters (ICAG) are performed at the Bureau International des Poids et Mésures (BIPM) in Sèvres and since 2003, with a 4-year time interval, also at the European Centre of Geodynamics and Seismology (ECGS) in Walferdange, Luxembourg. Such extensive comparison campaigns with a large number of absolute gravimeters may reveal biases not only between single instruments but also between different instrumental developments and technological realisations. Table 1.4 summarises the results from the comparisons ICAG89 (Boulanger et al. 1991), ICAG94 (Marson et al. 1995) and ICAG97 (Robertsson et al. 2001). In 1989, five JILA-type instruments and five individual developments participated. The JILAg-3 result differed from the mean of the JILA group by +1.8 μGal, from the mean of the group with individual developments by +3.3 μGal and in the average by +2.4 μGal from the mean of all 19 stations' determinations performed by the 10 instruments. In 1994, for the first time FG5 instruments contributed to the comparison, and the discrepancy of JILAg-3 to the mean result of all 11 meters was +2.8 μGal. These two comparisons may indicate a small offset of about +2 or 3 μGal for JILAg-3. In 1997, the situations changed somewhat. The sites A and A2 were observed, and for both points the JILAg-3 result was +5.5 μGal above the average of all instruments. In addition to these external comparisons with other gravimeters, the lower part of Table 1.4 shows an internal comparison for JILAg-3. Looking at the Clausthal series with respect to the whole time span (1986–2000), and the two periods 1986–1996 and 1997–2000, a systematic change in the measuring level cannot be detected. The Clausthal series neither confirms nor contradicts the ICAG97 experience. Both results are consistent considering the precision estimate of 4–5 μGal for a single station determination with JILAg-3.

From Table 1.4, it may be concluded that JILAg-3 was well embedded in the international absolute gravity definition. Overall, a larger discrepancy with other instrument groups did not really become obvious during the international comparisons. But a bias to the international standard, here defined as the average of all participating gravimeters at BIPM, of up to +5 μGal cannot be excluded. From the ICAG94 and ICAG97 comparisons, a measurement offset of +9 μGal becomes visible when just comparing JILAg-3 with FG5-101 as already mentioned. Thus, from the Hannover point of view, the offset correction for JILAg-3 has mainly to be considered as a bias with respect to the gravimeters FG5-220 and FG5-101 and not to the international standard. Interpreting the results of the international comparisons in Sèvres with respect to the instrument groups, a systematic error, inherent in the instrumental design of the JILAg or FG5 gravimeters, does not exist or is within the 1–2 μGal accuracy level. Nevertheless, temporary biases for single instruments are possible, e.g. due to undetected changes within the instrumental adjustments.

To investigate the stability of the presently employed gravimeter FG5-220 of IfE, Table 1.5 gives the results from the international comparisons in Walferdange (Luxembourg) 2003 and 2007 (external comparisons, Francis and van Dam 2006; Francis et al. 2010) and FG5-220 reference measurements in Bad Homburg (station of BKG, Wilmes and Falk 2006) from 2003 to 2008. Within 2 μGal, the Hannover FG5 instrument agrees with the internationally realised measuring level. With respect to the FG5-220 observations in Bad Homburg, it has to be mentioned

Table 1.4 JILAg-3 absolute gravity meter controlled by external (international) and internal (repetition) comparisons to ensure consistent long-term measurement accuracy (n: number of observations)

JILAg-3 external comparisons	Remarks	Gravimeter group	Mean g result [μGal]	Standard deviation of a single observation [μGal]	Δg [μGal] (JILAg-3 minus mean)
ICAG89, BIPM (Boulanger et al. 1991, Table 7)	Referred to *site A*, reference height 0.050 m, 19 stations' determinations with 10 absolute gravimeters	5 JILA	980,925,975.4	±6.2, $n = 11$	+1.8
		GABL, BIPM, IMGC, NIM, NAO	980,925,973.9	±9.2, $n = 8$	+3.3
		All 10 meters	980,925,974.8	±7.4, $n = 19$	*+2.4*
		Only JILAg-3	980,925,977.2	$n = 2$	
ICAG94, BIPM (Marson et al. 1995, Table 4)	Referred to *site A*0, reference height 0.900 m, 12 observations with 11 absolute gravimeters	4 JILA	980,925,710.3	±4.9, $n = 4$	+2.7
		6 FG5	980,925,710.4	±2.8, $n = 7$	+2.6
		1 IMGC	980,925,709.0	$n = 1$	+4.0
		All 11 meters	980,925,710.2	±3.3, $n = 12$	*+2.8*
		Only JILAg-3	980,925,713.0	$n = 1$	
ICAG97, BIPM (Robertsson et al. 2001, Table 5)	Occupied *site A* with 12 instruments, reference height 0.900 m	4 JILA	980,925,708.1	±5.5, $n = 4$	+5.6
		7 FG5	980,925,707.0	±3.7, $n = 7$	+6.6
		1 GABL-E	980,925,714.4	$n = 1$	−0.8
		All 12 meters	980,925,708.1	±4.5, $n = 12$	*+5.5*
		Only JILAg-3	980,925,713.6	$n = 1$	
ICAG97, BIPM (Robertsson et al. 2001, Table 5)	Occupied *site A*2 with 13 instruments, reference height 0.900 m	4 JILA	980,925,716.6	±3.5, $n = 4$	+3.5
		6 FG5	980,925,773.7	±2.9, $n = 6$	+6.4
		IMGC, NIM-2a, ZZB	980,925,713.9	±0.1, $n = 3$	+6.2
		All 13 meters	980,925,714.6	±5.0, $n = 13$	*+5.5*
		Only JILAg-3	980,925,720.1	$n = 1$	

Table 1.4 (continued)

JILAg-3 external comparisons	Remarks	Observation period	Mean g result [μGal]	Standard deviation of a single observation [μGal]	Δg [μGal]
Clausthal/Harz	IfE reference station for JILAg-3, 29 observations over 15 years, floor level	Period 1986–2000	981,115,734.5	±4.7, $n = 29$	
		Only 1986–1996	981,115,734.1	±4.8, $n = 20$	*−0.4*
		Only 1997–2000	981,115,735.4	±4.6, $n = 9$	*+0.9*

Table 1.5 FG5-220 absolute gravimeter controlled by external (international) and internal (repetition) comparisons to ensure consistent long-term measurement accuracy

FG5-220 external comparison	Remarks	Epoch	Δg [μGal] (FG5-220 – mean g)
ICAG2003, ECGS (Francis and van Dam 2006, Table 16)	13 absolute meters, 14 points, 52 determinations	November 2003	−1.9 standard deviation (single instrument) 1.8
ICAG2007, ECGS (Francis et al. 2010, Table 3)	19 absolute meters, 16 points, 73 determinations	November 2007	+2.4 standard deviation (single instrument) 2.0
FG5-220 internal comparison	**Remarks**	**Epoch**	**Δg (FG5-220) [μGal] (Single – mean g)**
Bad Homburg (gravimetry lab of BKG, Wilmes and Falk 2006)	Reference station for FG5-220 since 2003, point BA	February 2003	+0.9
		November 2003	−0.8
		April 2005	+1.2
		April 2006	+0.7
		November 2007	+0.2
		September 2008	−2.1

that the differences between the single epochs also contain real gravity changes due to time-varying environmental effects like seasonal hydrological variations. As shown in Table 1.5, the six stations' determinations agree very well, better than expected from empirical estimates, with a mean scatter of 1.1 μGal only (root-mean-square difference, r.m.s.). An instrumental instability cannot be identified. A similar experience is also gained from the yearly repetition surveys and from the comparisons with the other FG5 absolute gravimeters involved in the Nordic absolute gravity project, to determine the Fennoscandian land uplift, cf. Timmen et al. (2006b) and Bilker-Koivula et al. (2008).

1.4 Relative Gravimetry

The determination of gravity differences and variations requires a composite employment of absolute and relative instrumental techniques and observation methods. The optimal choice of the different types of available sensors allows one to organise the work in a most efficient way with respect to accuracy and economy. Relative gravimetry contributes among others to the following geodetic tasks:

- Support of absolute gravimetry (centring to safety points, national net points or to adjacent absolute points; measurement of vertical and horizontal gradients)
- Monitoring of temporal gravity changes in investigation areas with short transportation time spans between the measuring points
- Densification of national gravity reference networks
- Providing dense point data to improve regional geoids

Fig. 1.7 Scintrex Autograv CG-3 (*right*) and CG-5 (*left*) and a LaCoste–Romberg model G with carrying case (*in front*)

The accuracies one strives for are in the order of one to a few tens of microgals. For high-precision relative gravimetry, LaCoste–Romberg (LCR) spring gravimeters have been employed nearly exclusively over decades. For about 20 years, Scintrex has offered a different type of spring gravimeter, the Autograv CG-3, e.g. see Hugill (1988), and since 2003 the new CG-5. Figure 1.7 shows a LCR and two Scintrex meters. Based on the inventions of L. LaCoste and A. Romberg, the company ZLS Corporation, Austin, TX, USA, designed the automated Burris Gravity Meter (ZLS 2007). The instrumental investigations described in Jentzsch (2008) showed excellent results for the ZLS meter which may also be considered as a state-of-the-art instrument.

1.4.1 Principles of Spring Gravimeters

The principle of a vertical spring balance explains the general operation of a relative gravimeter. An elastic spring is used to generate a counterforce which keeps the sensor's test mass in equilibrium with the gravitational force, see left part in Fig. 1.8.

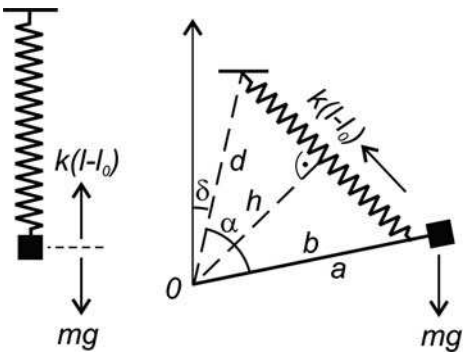


Fig. 1.8 The principle of a vertical spring balance (*left*) and of a lever spring balance (*right*)

For a translational system, which is in accordance with Hooke's law, the condition of equilibrium is given by

$$mg = k(l - l_0). \tag{17}$$

The spring constant k is the proportion of the stretching force, with mass m and gravity g, to the elongation (l–l_0) of the spring (l: spring length under load, l_0: length without load). To determine a gravity difference Δg, the small change in the spring length Δl has to be measured:

$$\Delta g = \frac{k}{m} \Delta l \, . \tag{18}$$

For that, a reading system with an extremely high resolution is required. To achieve a measuring precision of better than 10 μGal, the mechanical sensitivity of about 0.1 nm is needed with a corresponding time stability of the spring force. A precise knowledge of the calibration factor k/m can nowadays be obtained by measurements between well-known absolute gravity points.

The right part of Fig. 1.8 shows the general lever spring balance. The equilibrium condition for the torques generated by gravitational force mg and spring force $k(l–l_0)$ can be expressed with

$$mga \sin(\alpha + \delta) = k(l - l_0)h = k(l - l_0)b\frac{d}{l} \sin \alpha. \tag{19}$$

The equation shows a non-linear relation between gravity g and angle α. With the conditions

$$l_0 = 0 \;\; (\text{“zero-length spring”}) \quad \text{and} \quad \alpha + \delta = 90^\circ \, , \tag{20}$$

(19) simplifies to

$$mga = kbd \sin \alpha. \tag{21}$$

Choosing the angles

$$\alpha \to 90^\circ \quad \text{and} \quad \delta \to 0^\circ \tag{22}$$

increases the mechanical sensitivity considerably (“astatisation”).

The requirements in (20) and (22) are implemented in the design of LaCoste–Romberg gravimeters with a counterspring made of metal (Krieg 1981). To achieve a measuring precision of better than 10 μGal, a pick-off system with a resolution of a few hundred nanometers is needed only. Measurements with LCR meters require a very accurate alignment of the mass-beam part to the horizontal orientation. Due to a gravity change, the test mass diverges from the horizontal position which can be restored by turning a dial which moves the suspension point of the spring up or down (zero-method, “nulling” the beam). The whole transmission system consists of the

dial, a set of gear wheels, the measuring screw and a lever system. The difference between two readings of the dial, which is combined with a counter, corresponds to a gravity difference. In addition to this mechanical compensation for restoring the zero position, an electronic feedback system is used nowadays. The moveable middle part of a three-plate capacitor is attached to the test mass, which allows an electrical pick-off of the sensor position and a restoring to the zero position. The electronic feedback systems help to avoid periodical errors due to imperfections in the gear-screw construction (Schnüll et al. 1984; Röder et al. 1988).

With the technical advances in the 1980s, the company Scintrex, Concord, Ontario, Canada, was able to design a new relative gravimeter using the principle of the vertical spring balance. The Scintrex CG-3 and CG-5 gravimeters are non-astatised systems with a quartz spring which covers the worldwide gravity range and operates without any micrometer screw or gearbox. The capacitive transducer and electronic feedback system allows a pick-off resolution of 0.2 nm (Scintrex 1998). Besides the non-existence of periodic errors, an additional advantage of the linear spring system is that the sensitivity is independent of the inclination. The new Burris Gravity Meter from ZLS Corporation, which is based on the LCR astatisation principles with a lever spring balance, is equipped with a digital feedback system of about 50 mGal range to null the beam. The zero-length spring is made by a metal alloy and is characterised by its low drift (Jentzsch 2008).

For more details and a more extensive overview about the principles of relative gravimeters, refer to other available literature, e.g. Torge (1989).

1.4.2 Observation Equation

To transfer a raw gravimeter reading, here given in counter units, to a gravity value, the calibration has to be known. In addition, a time-dependent instability of the counterforce (fading of the spring tension) should be considered in the observation equation. Environmental disturbances, e.g. small temperature changes within the sensor case or mechanical impulses caused by transportation, may change the gravimeter's reading considerably. This instrumental drift can be modelled by a low-order polynomial and requires repeated measurements, temporally well distributed over the measuring period, of at least some of the network points. The following equation gives the connection between the raw reading and the resulting g-value for a LCR gravimeter:

$$g = N_0 + \sum_{j=1}^{p} d_j (t - t_0)^j + \sum_{k=1}^{m} Y_k z^k + \sum_{l=1}^{n} A_l \cos(\omega_l z - \varphi_l), \tag{23}$$

with N_0 = instrument level, d_j = drift parameter of degree j, t_0 =starting time (e.g. first daily measurement), Y_k = calibration coefficient of degree k, z = reading in counter units, A_l = amplitude, ω_l = frequency, ϕ_l = phase of the periodic term of degree l. Often the so-called Δg adjustment of a gravity network is applied,

Table 1.6 Maximum and mean amplitude of the periodic calibration terms as derived for 21 LaCoste–Romberg model G gravimeters in the gravimeter calibration system Hannover

Periods [CU]					
	Up to LCR-G457	1.00	7.88	35.47	70.94
	From LCR-G458	1.00	7.33	36.67	70.33
Maximum amplitude		8.1	15.2	21.5	18.0
Mean amplitude in [μGal]		2.5	5.2	5.9	8.0

which introduces gravity differences between two successive point measurements. The advantage is the elimination of the unknown N_0-parameter in the observation equation. For Scintrex CG meters, the periodic terms in (23) do not exist.

Table 1.6 summarises the magnitude of periodic errors for 21 LCR gravimeters as determined in the gravimeter calibration system Hannover. Neglecting these errors, an additional uncertainty (systematic error) of a few tens of microgals can be possible for gravity differences. Comparisons of the results for three LCR instruments from IfE, all three employed in the calibration systems Hannover and Wuhan/China (different gravity ranges), showed significant discrepancies for the polynomial and periodical calibration parameters (Xu et al. 1988). Therefore, for highly accurate measurements it is advised to examine the meter's calibration when transferring the parameters to different gravity ranges (recommendation from the author: for distances of more than 0.5 Gal away from the calibration system).

1.4.3 Regional and Local Surveys with Scintrex SC-4492

In 2001, IfE obtained the new Scintrex CG3 gravimeter no. 4492. The following investigations of this state-of-the-art relative gravimeter were focussed on the calibration (time stability and gravity range dependency). The study was performed over a time period of about 4 years and covered a gravity range of almost 1.5 Gal. In addition, other publications can be recommended to achieve a more general overview about the quality of the Scintrex Autograv CG-3 system, e.g. Hugill (1988), Jousset et al. (1995), Falk (1995), Rehren (1997) and Everaerts et al. (2002). With respect to instrumental precision, accuracy and drift, the IfE investigations confirm the results of the references given above.

Most of the surveys with the SG-4492 were carried out in the gravimeter calibration system Hannover; see Fig. 1.9. This was established between 1976 and 1982 for the determination of calibration functions for LCR gravimeters with 1 μGal accuracy (Kanngieser et al. 1983). The system serves for the analysis of polynomial and periodic calibration terms, with the intent of improving the manufacturer's calibration tables which usually provided accuracies of 10^{-3} to 10^{-4} only. Over 13,000 gravity differences measured with 47 LCR instruments and 12 absolute gravity determinations were included in the adjustment of the calibration system. The estimated mean standard deviations for the adjusted gravity differences are 2 μGal for the Cuxhaven–Harz line (~9 mGal intervals) and 1 μGal for the vertical calibration

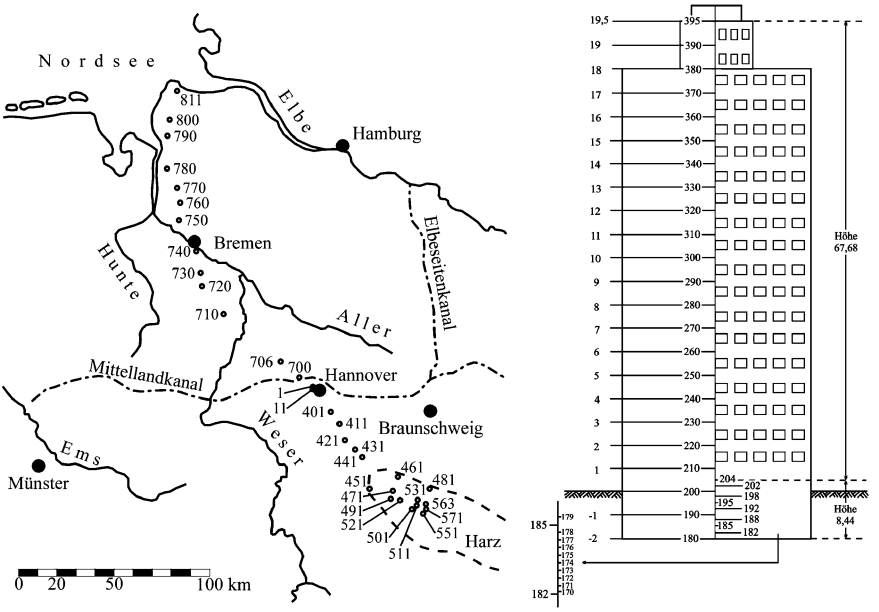


Fig. 1.9 The station distribution of the gravimeter calibration system Hannover (Cuxhaven–Harz Mountains, 300 mGal, and vertical calibration line Hannover, 20 mGal), cf. Torge (1989)

line Hannover (staircase of a 20-storied building, point intervals of 0.02, 0.2 and 1 mGal).

From November 2001 to September 2005, the SC-4492 has been employed in different projects in northern Germany and in Scandinavia; see Fig. 1.10. In most cases the instrument has been transported not only by hand but also by car. In general, the measurements were done using the step method to allow an optimal drift control, e.g. with a point sequence A-B-A-B-C-B-C-D-C-D. Each tie between neighbouring points was measured three times or more with a time span of 5–60 min between the two successive point occupations. Only during the absolute gravimetry campaigns in Fennoscandia in 2004 and 2005, the relative measurements between the absolute stations were observed once with a long time span of up to 10 h between the two successive readings. For each occupation three registrations with a read time (RT) of 60 s and a cycle time (CT) of 80 s were carried out. The seismic filter option of the online software was selected. The average of the second and third cycle was used for the post-processing with the program system GRAV from Wenzel (1993). The least-squares adjustment provides accuracy estimates for the single gravity difference observations in the order of 4–10 μGal when excluding the two long-distance campaigns in Fennoscandia. Measuring gravity ties with short transportation ways, points can be connected within an accuracy level of about 1 μGal. The points in the Harz Mountains (461–571) show small discrepancies between the calibration line reference values and the recent results from SC-4492. The differences vary between −3.8 and +3.1 μGal with an r.m.s. discrepancy of 2.4 μGal.

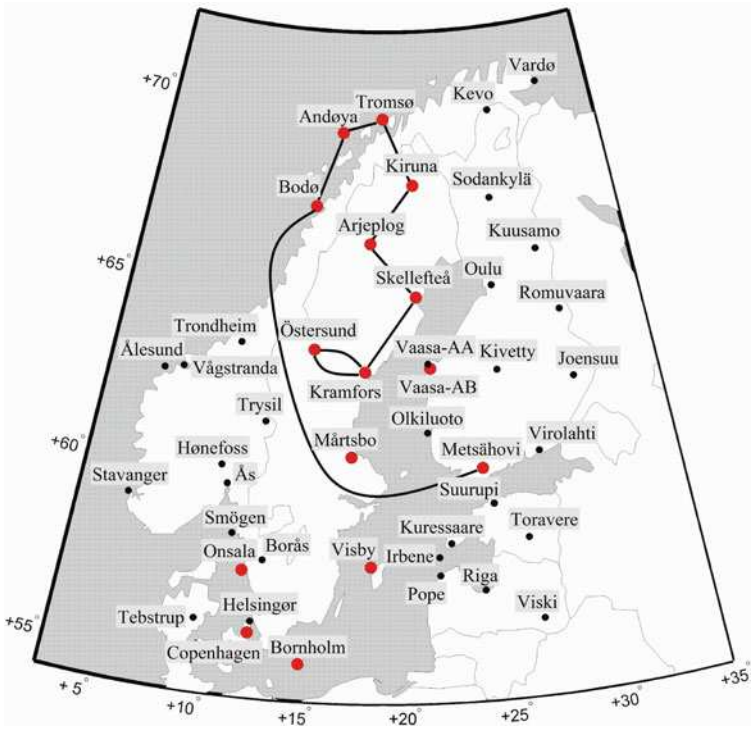


Fig. 1.10 Absolute gravity stations in the Fennoscandian land uplift area occupied by FG5-220 in 2004 (*red dots*). Lines show the ties used to control the calibration of the relative gravimeter SC-4492

For the northern part of the calibration line, the differences are much larger (maximum +6.5 μGal and minimum −11.4 μGal, r.m.s. 6.8 μGal). Besides measurement errors, the large discrepancies can also be caused by groundwater and soil moisture effects and/or secular gravity changes during the last 20 years.

Table 1.7 summarises the calibration results obtained for SC-4492, standing as an example for a present-day advanced relative gravimeter. Polynomial calibration terms of higher degree were not found. The calibration factors E (improvement of the manufacturer calibration) were obtained with a precision between 2 and 8×10^{-5} and are varying within a range of 3×10^{-4}. Calculating a mean factor and expressing the single deviations from the mean in gravity discrepancies (last column), disagreements up to almost 20 μGal were found. But these values cannot be assigned to instabilities or gravity range dependencies of the calibration factor. The uncertainties in the reference gravity values and, moreover, the subsurface water mass changes (groundwater, soil moisture, crevasses and clefts in rock filled with water) can introduce errors of more than 10 μGal. Therefore, a time-dependent instability of the calibration in the order of 1×10^{-4} cannot be excluded but is also not proven. In addition, the calibration results show no correlation with the different

Table 1.7 Calibration of Scintrex CG-3 no. 4492 in different gravity ranges over 4 years (adjustment results: calibration factor, accuracy, drift), total gravity range: 1,468 mGal

Project	Date	Δg (max) [mGal]	$n\ \Delta g$	E	Drift [μGal/h]	$s\ (\Delta g)$ [μGal]	ΔE [$\times 10^{-6}$]	$\Delta E \rightarrow \delta g$ [μGal]
Calibration line Hannover–Cuxhaven	23.11.2001	118	4	1.000854± 0.000026	18.7	±13.5	65	+7.7
Connection of two absolute stations Hannover–Harz	04.07.2002	147	4	1.000912± 0.000034	24.8	±10.2	7	+1.0
Calibratione line Hannover–Harz	05.11.2002–22.01.2003	190	104	1.001011± 0.000025	18.0	±10.5	−92	−17.5
Connection of two absolute stations Hannover–Harz with intermediate points	14.11.2002–28.11.2002	147	42	1.001031± 0.000031	16.9	±5.4	−112	−16.5
Calibration line Hannover–Cuxhaven	23.01.2003–15.02.2003	117	127	1.001062± 0.000060	11.7	±9.9	−143	−16.7
Connection of two absolute stations Hannover–Hamburg	27.02.2003	108	13	1.000786± 0.000020	19.6	±3.7	133	+14.4
Calibration line Harz mountains	21.04.2004	93	14	1.000872± 0.000026	32.7	±8.1	47	+4.4
Fennoscandian land uplift area 2004	02.06.2004–02.07.2004	623	16	1.000945± 0.000076	13.0	±24.7	−26	−16.2
Calibration line Harz mountains	13.04.2005	93	11	1.000836± 0.000027	9.8	±7.4	83	+7.7
Fennoscandian land uplift area 2005	20.08.2005–25.09.2005	445	12	1.000878± 0.000042	11.4	±23.6	41	+18.3
Mean				1.000919± 0.000029	17.7 (≡0.43 mGal/day)			

$n\ \Delta g$: number of Δg measurements; E: adjusted linear calibration factor; drift: adjusted linear drift factor; $s\ (\Delta g)$: standard deviation of a single Δg measurement; ΔE: deviation from mean E; $\delta g = \Delta E \times \Delta g$ (max)

gravity ranges which lead to the conclusion that no gravity range dependence exists over the total investigation range of 1,470 mGal.

1.4.4 Microgravimetric Measurements

Highest accuracy can be expected for measurements in a small network with points distributed in one room or in a single building (short time spans between measurements, meter transportation manually (shock prevention), no wind, stable temperature). In an extensive survey, the vertical calibration line in Hannover has been measured with SC-4492 (31 points, 328 gravity difference observations), cf. Table 1.8. The standard deviation for a single gravity difference measurement is just 3.7 μGal. Figure 1.11 reveals a systematic discrepancy between the calibration line reference values and the new determined figures. The differences for the points above ground floor show a height and gravity dependence which can be interpreted as a linear scale error of about 3×10^{-4} ($\equiv$6.0 μGal). After these investigations with SC-4492, it cannot be excluded that the vertical calibration line is deteriorated by a small-scale error. Additional investigations with another CG-3 or CG-5 are needed to clarify this issue. One reason for the discrepancies of points below ground floor may be due to the different gravimeter setups. The LCR meter has normally an average sensor height of about 6 cm above floor level. The CG-3 system with its tripod measures the gravity at a height of about 27 cm. The points are all in corners very close to the walls, only 20 cm away. The not well-known gravity gradients along the vertical, with their non-constant behaviour, disturb the comparison of the different gravimeter systems. The r.m.s. discrepancy between the recent SC-4492 results and the reference values is 2.3 μGal.

The determination of vertical gravity gradients is important, because the combination of instruments with different reference heights strongly needs a highly precise centring of the measured gravity values to a common reference. Vertical gradients were observed at the two absolute gravity stations of IfE, Hannover and Clausthal, and at stations of the Fennoscandian uplift area, cf. Tables 1.8 and 1.9. With the help of a tripod of 1 m height, the gravity difference is measured to determine the gradient, cf. Fig. 1.5. Because of the sensor height difference between the LCR and the CG-3 systems (about 21 cm), the results from the two kinds of meters can differ by some 1 μGal. The LCR meters of IfE are equipped with the SRW-feedback systems, which eliminates the problems with periodic errors and gravity dependencies for small gravity differences (Röder et al. 1985). Figure 1.12 shows the absolute gravity meter setup on the pier in the basement of station Clausthal. A mesh of nine points with a spacing of 40 cm has been surveyed with SC-4492 to determine the horizontal gravity field above the pier surface. The result (Fig. 1.12, right) seems to be reasonable. With distance to the wall (cf. Fig. 1.12, left), gravity increases by about 2.5 μGal per 10 cm.

The obtained accuracies for all microgravimetric surveys are in the order of 1–2 μGal. In Table 1.9, three vertical gradients are compared with LCR results. In all cases the obtained results from SC-4492 are smaller than the LCR results which

Table 1.8 Network and adjustment statistics/results (accuracy, drift, etc.): re-measurements of 31 points of the vertical calibration line Hannover, determination of gravity gradients with Scintrex CG-3 no. 4492

Project (indoor observations)	Date	Δg (max) [mGal]	Number of points	$n\ \Delta g$	Δt [min]	Drift [μGal/h]	s (Δg) [μGal]	s (point) [μGal]
Vertical gravity calibration line Hannover	18.05.2002–6.08.2002	19.2	31	328	~7–10	25.1	±3.7	±1.2
Vertical gravity gradient at absolute station Hannover	05.12.2002	0.30	2	20	~5–6	7.4	±2.4	±0.5
Vertical gravity gradient at absolute station Clausthal/Harz	30.01.2003	0.27	2	20	~3	0.4	±4.2	±0.9
Horizontal gravity gradient at absolute station Clausthal	30.01.2003	0.02	9	47	~3	4.7	±4.6	±1.4
Eleven vertical gravity gradients in the Fennoscandia uplift area	August/September 2003	0.38	11 × 2	11 × 10	~5–6			±1.7

$n\ \Delta g$: number of Δg measurements; Δt: time span between successive observations; drift: adjusted linear drift factor; s (Δg): standard deviation of a single Δg measurement; s (point): standard deviation (mean) for a gravity point.

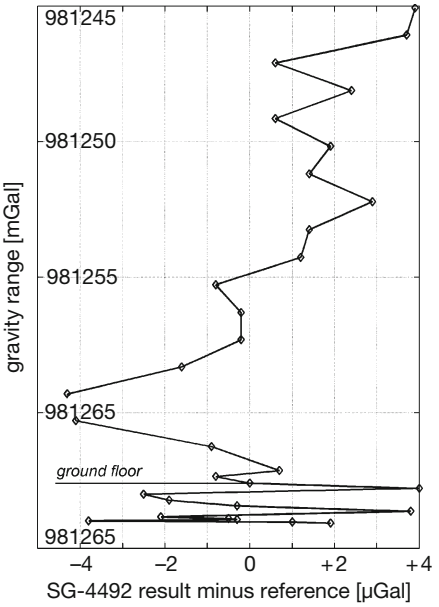


Fig. 1.11 Differences between the reference values of the vertical calibration line and the new determined gravity values with SC-4492

Table 1.9 Comparison of the SC-4492 results with reference results (all determined with LCR gravimeters)

Project	Comparative figures	Difference to SC-4492
Vertical gravity calibration line Hannover	Calibration line reference values	±2.3 μGal (r.m.s.)
Vertical gravity gradient at absolute station Hannover	Five LCR SRW-feedback meters in 1993/1994, mean: 303.1 μGal/m	+1.8 μGal/m
Vertical gravity gradient at absolute station Clausthal/Harz	Four LCR SRW-feedback gravimeters in 1987, mean: 266.0 μGal/m	+3.0 μGal/m
Vertical gravity gradient at absolute station Vaasa (AB)	Simultaneous observation with LCR-G709 SRW-feedback gravimeter in 2003: 330.7 μGal/m	+2.8 μGal/m

are reasonable for these stations considering the different sensor heights above the massive concrete piers.

1.4.5 Instrumental Drift

The gravimeter drift can be differentiated into two parts: stationary drift mainly due to spring aging and the transportation drift which may be caused by small shocks, vibrations, temperature effects or hysteresis effects after sudden changes of

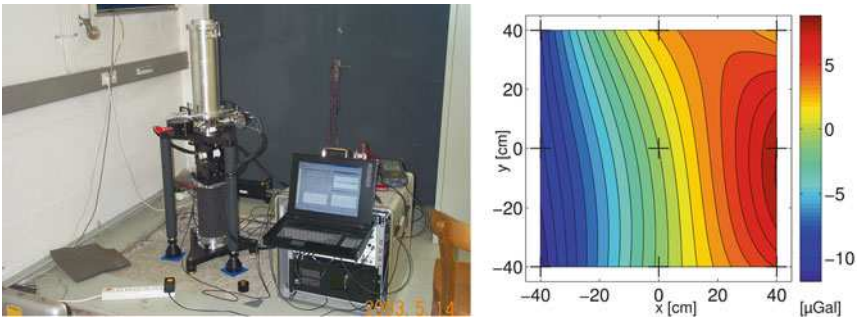


Fig. 1.12 The FG5-220 occupying the absolute gravity point at station Clausthal; contour plot of the horizontal gradient field above the pier with 1 µGal intervals

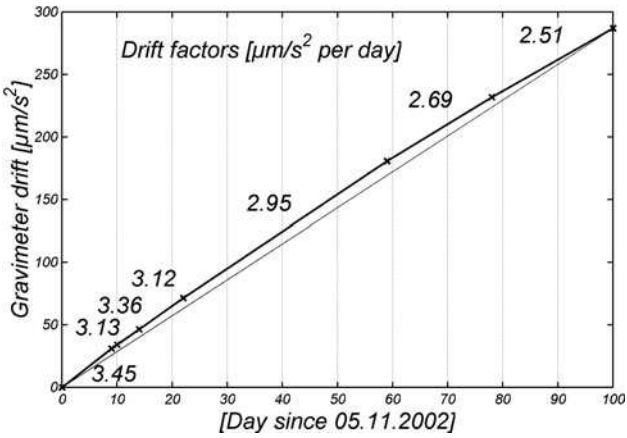


Fig. 1.13 Long-term drift (composition of stationary and transport drift) of SC-4492

the spring load (e.g. mechanical unclamping of the lever). A long-term drift (composition of stationary and transportation drift) of the meter is shown in Fig. 1.13. On 8 different days within a time span of 100 days, the first reading in the morning on a common starting point has been used to derive this long-term behaviour. The figure depicts a nearly linear behaviour.

Table 1.10 summarises the adjusted linear drift factors from daily field surveys obtained on 14 different days. It becomes clear that the drift behaviour of SC-4492 during the field surveys is significantly not linear. The drift can vary enormously. Therefore, for precise geodetic measurements a drift behaviour has to be taken into account as a non-linear temporal change of the zero level of the gravimeter's readings. The drift is determined by repeated point occupations during a day which allows a modelling by a low-order polynomial with time; see (23). Depending on the network structure, the instrumental behaviour and the required accuracy, different measurement methods can be applied to control and determine the drift of

Table 1.10 Adjusted linear daily drift factors from daily field surveys with SC-4492

Date	Number of points	Number of Δg observations	Measuring time [h]	Drift [μGal/d]
05.11.02	11	14	8	436
14.11.02	5	12	9	361
15.11.02	7	19	6	532
20.11.02	2	10	2	433
28.11.02	3	10	6	383
05.12.02	2	20	2	177
04.01.03	2	9	5	309
21.01.03	4	6	4	378
22.01.03	2	4	2	592
23.01.03	4	5	6	69
05.02.03	3	6	6.5	78
12.02.03	4	9	6.5	−49
15.02.03	2	10	3	412
27.02.03	2	3	6.5	469
			Mean	327±189

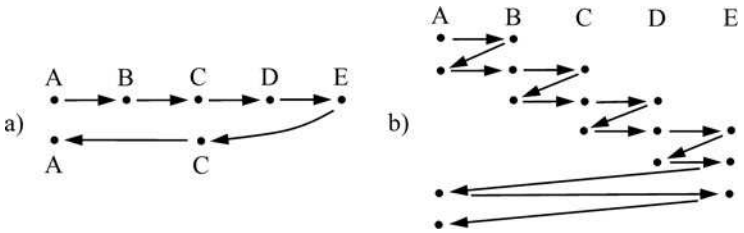


Fig. 1.14 Measuring procedures for drift control: (**a**) profile method with weak drift control, (**b**) step method with strong drift control (three times each tie and overlaps)

an instrument which is shown in Fig. 1.14. In general, the step method is used for most precise surveys. For example, the Hannover gravimetry group applies the five times step method to calibrate gravimeters with electronic feedback systems using the vertical calibration line in the 20-storied university building.

1.5 Reduction of Non-tectonic Gravity Variations

The Earth's gravity field varies continuously with time which is explained in detail, e.g. by Torge (1989). For geodynamics research, the establishment of regional gravity control networks and the establishment of globally distribute absolute gravity stations serve to reveal gravity changes of long-term or secular character. Such changes may occur together with tectonic plate movements, with postglacial isostatic compensation processes, with tectonic processes like mountain building or with compactions in sediment basins. Local gravity changes of short-term or even

abrupt character may become detected by gravity monitoring nets covering areas with volcanism or earthquakes. In addition, human activities can cause significant variations in the Earth's gravity field (large constructions, withdrawal of water, oil, gas, etc.). All gravimetric measurements are subject to irregular and periodic changes caused by tides, groundwater and other hydrological processes (e.g. soil moisture variations), atmospheric mass movements and polar motion. These effects of non-tectonic causes are superimposed on the target signal and have to be removed as well as possible. Generally, gravimetric measurements are freed from effects of the tides, the atmospheric mass redistributions and the small movements of the Earth's rotation axis within the Earth. The transfer function between changes in the groundwater table and the related gravity effect at the measurement point is often not well known, and therefore the latter is not a standard reduction in gravimetry. Torge et al. (1987) describe the reductions for absolute gravity measurements with the Hannover JILA absolute meter.

1.5.1 Earth's Body and Ocean Tides

The tidal deformation of the Earth is an elastic response of its body to the gravitational accelerations produced by the Moon, the Sun and, to a slight extent, also by the planets (Wang 1997; Wenzel 1997). At mid-latitudes periodical deformations over a day occur with an amplitude range of up to 40 cm. The maximal gravity variation remains below 300 μGal. The solar tides amount to 46% of the lunar tides.

In the Earth's centre of mass, the gravitation of the other celestial bodies is completely compensated by the centrifugal accelerations due to the orbital motion of the Earth. Figure 1.15 shows a simplified version of the Earth–Moon system with a non-rotating Earth and the Moon as a point mass. The motion of the Earth around the barycentre of the two-body system generates the orbital acceleration $-\boldsymbol{b_0}$ which is a constant for all points within the body and on the surface of the Earth. Because of the spatial extension of the Earth, the gravitation vector $\boldsymbol{b}$ differs from position to position. The tidal acceleration $\boldsymbol{b_t}$ for point P on the Earth's surface is the sum of the gravitation $\boldsymbol{b}$ and the orbit acceleration $-\boldsymbol{b_0}$. Applying Newton's gravitational law, the tidal acceleration vector for the Moon is given by

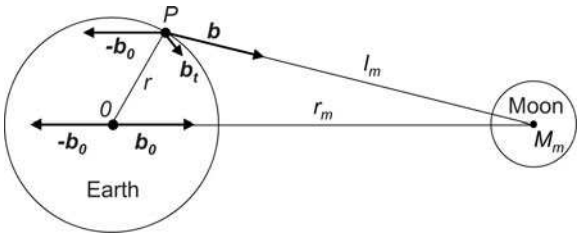


Fig. 1.15 Tidal acceleration as the sum of the Earth's orbital acceleration and lunar gravitation

$$b_t = b - b_0 = \frac{GM_m}{l_m^2}\frac{\mathbf{l_m}}{l_m} - \frac{GM_m}{r_m^2}\frac{\mathbf{r_m}}{r_m}, \tag{24}$$

where M_m is the mass of the Moon, G the gravitational constant, l_m and r_m the distances between point P and the Moon and between the Earth's centre of gravity 0 and the Moon, respectively.

Because of the Earth's rotation and the continuously varying distances of the Sun and the Moon from the Earth, a large number of waves (partial tides) have to be considered to model the theoretical tides for a rigid (not deformable) Earth's body. The ephemerides of the celestial bodies are well known from astronomy which allows the precise calculation of the rigid Earth tides. The tidal spectrum comprises a long-periodic part with half-monthly, monthly, semi-annual, annual and longer periods and short-periodic waves with the main power in the daily and half-daily periods. The longest wave with 20,942 years is the period of the mean ecliptic longitude of the sun's perigee. The widely available harmonic developments of the tidal potential follow the spectral representation as chosen by Doodson (1921). In the more recent literature, a detailed mathematical description can be found, among others, in Wenzel (1997). Tidal potential catalogues with different accuracies contain up to more than 1,000 waves. The most common catalogues are from Cartwright, Taylor and Edden (Cartwright and Taylor 1971; Cartwright and Edden 1973) with 505 partial tides (accuracy better than 0.1 μGal) and from Tamura (1987) with 1,200 waves (0.01 μGal). The model from Hartmann and Wenzel (1995) with 12,935 waves also includes coefficients due to the nearby planets and to the flattening of the Earth. An extensive description of the principal waves of the theoretical tides (rigid Earth) is given in Zürn and Wilhelm (1984).

Gravimetric Earth tide measurements show large differences with respect to the theoretical tides which can be explained by the non-rigid behaviour of the Earth's body and by effects from the ocean tides. The astronomical tide generating forces cause an elastic deformation of the solid Earth. Compared to the model of the rigid Earth, the amplitudes of the partial tides of the solid Earth are amplified and a phase shift takes place. In addition, the ocean tides affect the gravity measurements by the direct attraction of the moving water masses and indirectly by the resulting deformation of the crust due to the water load (ocean load tides). In general, close to the ocean the tidal loading effect is much smaller than the body tides but still affects gravity to very large distances from the coast (Jentzsch 1997). The ocean loading signal is not in phase with the body tides.

To reduce gravimetric measurements for Earth's body and ocean tides, the gravimetric tidal reduction as a compound tidal signal can be described as a sum of periodic terms:

$$\Delta g_{\text{tid}} = -\sum_{i=1}^{n} \delta_i A_i(\text{theor}) \cos\left(\omega_i t + \Phi_i(\text{theor}) + \Delta\Phi_i\right) \tag{25}$$

$$\text{with amplitude factor} \quad \delta_i = A_i \div A_i(\text{theor}) \tag{26}$$

Table 1.11 Principal gravimetric partial tides and the measured amplitude factors and phase leads (station Hannover: $\varphi = 52.389°$, $\lambda = 9.714°$). The Earth tide registrations with spring gravimeters in Hannover did not allow the evaluation of long-periodic tidal waves due to the instrumental drift behaviour. The partial tide with 18.6 years (6,798.4 days) has no special symbol

Tide symbol	Period [solar days/h]	Amplitude (theor) [μGal]	Amplitude factor δ [ ]	Phase lead $\Delta\Phi$ [°]
Constant lunar and solar tides				
M_0S_0	∞	26.7	–	–
Long-period tides				
–	6,798.4 days	2.4	–	–
Sa	365.26 days	0.4	–	–
Ssa	182.62 days	2.6	–	–
Mm	27.55 days	2.9	–	–
Mf	13.66 days	5.6	–	–
Diurnal tides				
Q_1	26.87 h	5.8	1.151	0.05
O_1	25.82 h	30.0	1.150	0.15
P_1	24.07 h	14.0	1.149	0.11
K_1	23.93 h	42.3	1.140	0.13
Semidiurnal tides				
N_2	12.66 h	5.4	1.176	2.34
M_2	12.42 h	28.0	1.186	1.68
S_2	12.00 h	13.0	1.189	0.41
K_2	11.97 h	3.5	1.191	1.07
Terdiurnal tides				
M_3	8.28 h	0.3	1.068	0.47

$$\text{and phase lead} \quad \Delta\Phi_i = \Phi_i - \Phi_i(\text{theor}). \tag{27}$$

The frequencies of the partial tides $\omega_i = 2\pi/T_i$ (T: period), the amplitudes A_i(theor) and the phases Φ_i(theor) are already derived from the models for the theoretical tides. The amplitude factor δ is also called gravimetric factor and can be determined together with the phase lead $\Delta\Phi$ by comparing the results of a continuously recording relative gravimeter with the rigid (theoretical) Earth tides. Table 1.11 summarises the results of Earth tide registrations in Hannover with five LCR gravimeters equipped with SRW-feedback systems (Timmen and Wenzel 1994). In the past, the global factor $\delta = 1.16$ was often used for all tidal waves in case of not existing observed parameters.

In contrast to tidal observations, the gravimetric tides can be computed on the basis of a model of the Earth's body determined from seismology, e.g. Preliminary Reference Earth Model (PREM) from Dziewonski and Anderson (1981), and using a global ocean model derived from satellite altimetry or from tide gauge observations as done by Schwiderski (1980). The latter used tidal observations of tide gauges along the continental coasts and on islands and developed a hydrographical interpolation model. For the first time, a global model was available describing the tidal response of the ocean's water masses. With the successful satellite altimetry

missions like Geosat (1985–1990), TOPEX/Poseidon (1992–2006) and some others, direct measurements of the ocean's surface height were evaluated to provide accurate ocean tide models. Because Schwiderski's model is still accurate enough, it is widely used for the tidal reduction in absolute and relative gravimetry.

At IfE in Hannover, the series development from Tamura (1987) delivers the tidal effects for the solid Earth, with synthetic tidal parameters (amplitude factors and phase shifts) interpolated from a worldwide l° × l° grid (Timmen and Wenzel 1995) to take the Earth's elastic behaviour into account. This grid was computed from

- Body tide amplitude factors using the Wahr–Dehant model (Wahr 1981; Dehant 1987) of an ocean-free, uniformly rotating and ellipsoidal Earth with inelastic mantle, liquid outer core and elastic inner core
- Ocean tide gravitation and load (Agnew 1997) derived from a l° × l° ocean tide model (Schwiderski 1980)

For the time-constant M_0S_0 tides, the amplitude factor 1.000 and a zero phase shift are used according to the IAG standards (Rapp 1983, "zero-tidal gravity"). For absolute gravity measurements the uncertainties in the geographical coordinates should be less than 10" with a height accuracy of better than 100 m. The time of a gravity observation can easily be recorded with better than 10 s UTC. Because the measurements of a station determination are distributed over 1–3 days, the average result can only be affected by residual errors of some 0.1 μGal (Timmen 1994). Near the coasts, larger uncertainties are possible.

1.5.2 Polar Motion

The Earth's rotation vector varies its orientation with respect to the Earth's crust. The penetration points of the rotation vector through the Earth's surface, the poles, are subject to motions of several metres per year. Figure 1.16 depicts the winding curve of the instantaneous North Pole relative to the reference pole of the International Earth Rotation and Reference Systems Service (IERS). The plotted x_p and y_p pole coordinates are provided by IERS on their internet pages. They are defined in a plane tangential to the pole with the x-axis in the direction of the Greenwich mean meridian and the y-axis points to the 90°W meridian. The polar motion consists mainly of two periodic components and a long-term irregular drift (cf. Figs. 1.16 and 1.17):

- The Chandler period (wobble) of 435 days has an amplitude of about 0.1" to 0.2" (3–6 m). This free oscillation is due to the dynamical flattening of the Earth and is excited when the instantaneous rotations axis deviates from the principal axis of inertia (figure axis). The mass displacements in the atmosphere and in the oceans exert torques on the solid Earth and excite the Chandler wobble continuously. An

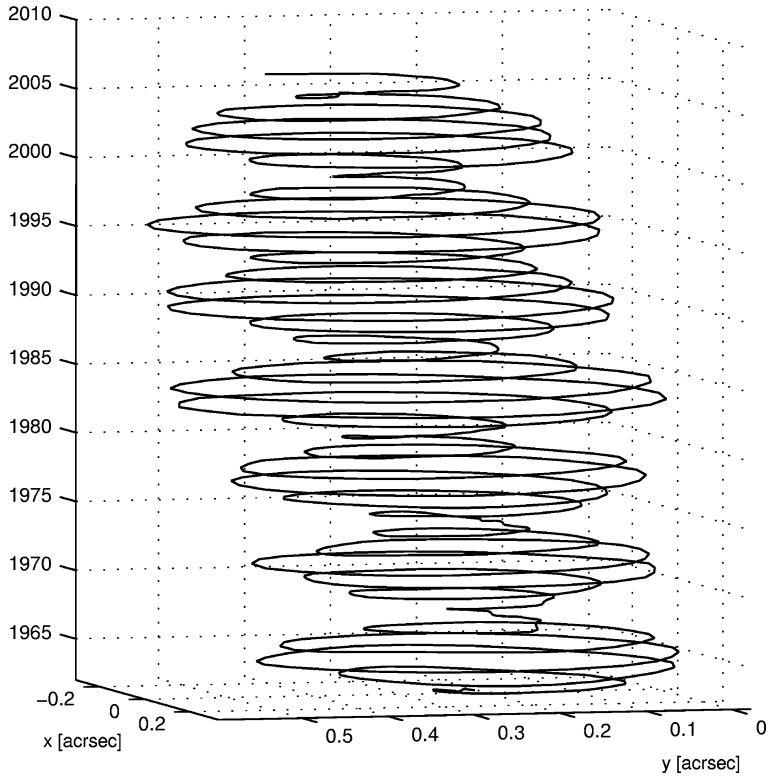


Fig. 1.16 Polar motion plotted with the IERS pole coordinates from 1962 to 2008 as published by the International Earth Rotation and Reference Systems Service on the internet

effect of earthquakes on the Chandler wobble is discussed in literature but is not proven up to now.

- An annual period is superimposed on the Chandler wobble with amplitudes of 0.05" to 0.1" and is caused by seasonal variations in the atmosphere and in the oceans.
- A secular motion of the North Pole is directed to the 70°W meridian with a magnitude of about 0.003" (0.1 m) per year (Fig. 1.17). The postglacial land uplifts in northern Canada and Europe are assumed to be the main causes for the pole drift presently. In addition, the lithosphere plates move horizontally against each other on the less viscous asthenosphere (plate tectonics) which appears as a pole drift (polar wander).

The interaction between Earth rotation and global geodynamical processes is comprehensively explained in Schuh et al. (2003). The superposition of Chandler and annual period induces a modulation (beating) period of 6 years which is clearly seen in Fig. 1.16.

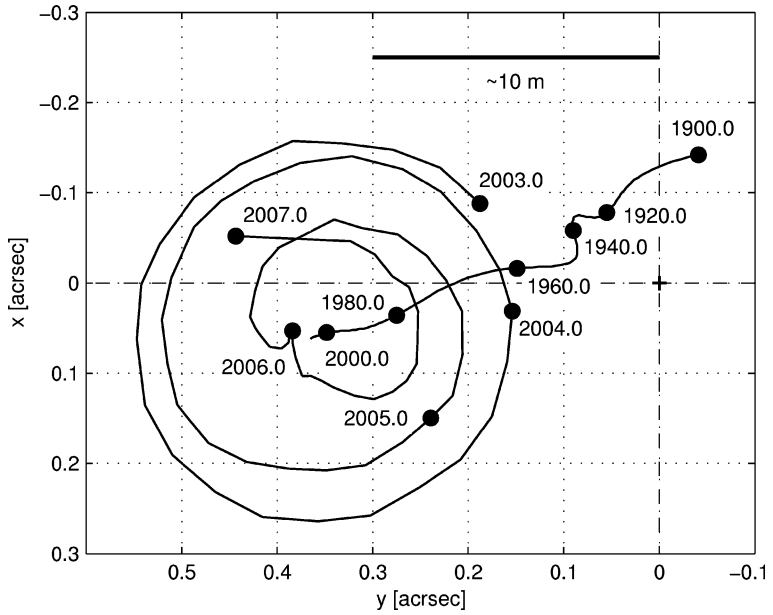


Fig. 1.17 Polar motion 2003–2007 and mean yearly motion calculated for the time points 1,900.0–2,000.0 as viewed from the North Pole (pole coordinates provided by IERS)

The variations of the Earth's rotation vector change the centrifugal acceleration at any measuring point on the Earth's surface. For station Hannover ($\phi = 52.44°$N, $\lambda = 9.71°$E) the gravity effects are normally within $\pm 5\,\mu$Gal, but attained a maximum of +7.3 and a minimum of $-8.6\,\mu$Gal around 1950. From Fig. 1.17 it becomes obvious that nowadays and in the future the polar motion effect for stations located along the positive direction of the *y*-axis (North and South America) would be obtained with a positive sign.

The polar motion reduction (Wahr 1985) for absolute gravimetry measurements are given as

$$\Delta g_{\mathrm{pol}} = -\delta_{\mathrm{pol}}\,\omega^2 r\,\sin 2\varphi\,(x_{\mathrm{p}}\cos\lambda - y_{\mathrm{p}}\,\sin\lambda)\,, \tag{28}$$

with the Earth's angular velocity ω and radius r, and the geographical latitude φ and longitude λ of the measuring position. The amplitude factor δ_{pol} considers the elastic response of the solid Earth as compared to a rigid Earth's body. As for the Earth tide amplitudes in the past, the factor 1.16 is applied here for the lack of better knowledge. The daily pole coordinates x_{p} and y_{p} of IERS (Bulletin A) are provided as predicted values, which can be used during the online data evaluation of absolute gravity measurements, and as finals which improve the gravity results by post-processing. The high accuracy of the IERS coordinates (± 0.0003", Reigber and Feissel 1997) keeps the residual error of this reduction below 0.1 μGal (Timmen

1994). The position coordinates are required with an accuracy of a few kilometres only.

1.5.3 Atmospheric Mass Movements

Gravity variations due to atmospheric mass variations may be subdivided into a direct effect of air mass attraction and an indirect (loading) effect due to the deformation of the Earth's crust and the sea surface. The surface deformation at the gravimeter site can be more than 1 cm. The variations in the local gravity acceleration and atmospheric pressure are known to be correlated with an admittance of -0.3 to $-0.4\,\mu$Gal/hPa depending on the local, regional and global weather situation (air mass distribution) and on the location of the gravimetry station, e.g. vicinity of the sea. During relative and absolute gravity measurements, the local air pressure is observed to reduce the atmospheric effect from the gravity results. In accordance with the IAG resolution No. 9, 1983 (IGC 1988), the factor $\alpha = -0.3\,\mu$Gal/hPa should be applied as a global mean if no better information is available. The reduction formula

$$\Delta g_{\text{air}} = -\alpha\,(p - p_{\text{n}})[\text{hPa}] \tag{29}$$

refers the actual atmospheric pressure p to the normal atmospheric pressure p_{n} which is defined by

$$p_{\text{n}} = 1013.25\left(1 - \frac{0.0065\,H_{[m]}}{288.15}\right)^{5.2559}\ [\text{hPa}]\ . \tag{30}$$

The reduction applies the US Standard Atmosphere, 1976, as a reference atmospheric model. The station height H (above sea level) should be introduced with an accuracy of better than 10 m for precise absolute gravity determinations.

Absolute gravimetry is more sensitive to atmospheric variations than relative gravimetry because of the short time intervals between two successive relative readings of a gravity difference. Assuming an air pressure variation of more than 30 hPa, the actual coefficient α should be known with an accuracy of better than 5% to ensure a reduction uncertainty of less than 0.5 μGal. But that requirement can often not be fulfilled especially when just using the general factor $-0.3\,\mu$Gal/hPa.

At IfE, a more accurate reduction is applied for all FG5 measurements performed in the Fennoscandian land uplift project since 2003. The attraction and deformation effects for a local (spherical distance $\leq 0.5°$), regional (0.5–10°) and global (10–180°) zone with corresponding resolutions of 0.005°, 0.1° and 1.125° are calculated. The global data are available from the European Centre for Medium-Range Weather Forecasts (ECMWF) and are provided to IfE by the University of Cologne in cooperation with the German Computing Centre for Climate and Earth System Research. The calculation procedure is explained in Gitlein and Timmen (2006).

This improved reduction for absolute gravity measurements with global atmospheric weather data is very costly in terms of labour and needs the development of gravimetric software which is not available in the open market. An improvement of 0.5–1 μGal can be expected (Gitlein 2009) but a residual error in the order of 0.5 μGal can still not be excluded for all gravimetric station determination.

1.5.4 Groundwater Variations

Gravity changes caused by groundwater variations are predominantly a direct gravitational effect of the water masses. Smaller indirect effects are accompanied with a vertical displacement of the measuring position: elastic deformation of the crust due to the water load and sediment consolidation due to a decrease in the water table level and a consequent decline of the pore volume (Romagnoli et al. 2003). Such effects can partly be avoided by selecting stations in mountain areas and bedrock. Temporary water storage is still possible in clefts, crevasses and pockets but the void volume should be much less than 5%, whereas in glacial sediment layers the free volume might be more than 30%. In general, a seasonal behaviour of the groundwater table becomes visible in continuous registrations of water depth gauges. Hence it might be helpful to perform absolute gravity measurements always in the same season of a year to determine secular gravity changes.

For regions with homogeneous sediment layers, a Bouguer-plate model often gives a first approximation for the dependency between changes in gravity and water table readings (Torge 1993):

$$\Delta g_{gw} = 2\,\pi\, G\, \rho_w\, P\, \delta H \;=\; 42\; P\, \delta H_{[m]} \quad [\mu\text{Gal}], \tag{31}$$

where G is the gravitational constant and ρ_w is the water density which is 1,000 kg/m^3. Assuming a pore volume of 30% in the sediment layer ($P = 0.30$) and a water table shift δH of 1 m, a gravity effect of 12.6 μGal is obtained.

For the absolute reference station Hannover, a vertical sediment profile was determined from the drilling of the water gauge close to the gravimetry laboratory. It revealed an average pore volume of 38%. Modelling the change of the groundwater table with a tilted plane (0.4% to the receiving river) and a mean water level depth of 3.7 m, an admittance factor of 15 μGal/m was derived (Timmen 1994). A similar relation is obtained from the statistical correlation of the water table readings with the absolute gravity measurements performed with the Hannover FG5 meter since 2003 (Sect. 1.6.1). The resulting regression factor with 17 μGal/m considers not only the primarily affecting Newtonian attraction but also the indirect effects accompanied with a vertical position shift (load and sediment consolidation).

Successful experiments and modelling of gravity effects due to soil moisture and groundwater changes were already described in Mäkinen and Tattari (1988). But, in general, a reduction for hydrological variations is still not applied in absolute and relative gravimetry.

1.6 Gravity Changes: Examples

Some examples for gravimetric applications of IfE are given in Torge (1993) describing projects in tectonically active areas in northern Iceland, the Venezuelan Andes and in the Yunnan (China) earthquake study area. Gitlein et al. (2008) describe the gravimetric survey of the Fennoscandian postglacial rebound which is an isostatic uplift of the Earth's crust due to the melting of the ice sheet after the glacial maximum of the last ice age. The IfE gravimetry projects in the Nordic countries are all part of a long-term survey task. In close cooperation with the national Nordic surveying agencies and research institutions, IfE has performed gravity observations in Scandinavia since 1986. The monitoring of crustal deformations in northern Europe is still the main focus of the ongoing cooperations.

As already mentioned in Sect. 1.2.2, an important future application may be the monitoring of changes in the hydrosphere, especially if such variations mean some serious consequences for the water supply. To demonstrate the potential of gravimetry as a tool for groundwater monitoring, the situation in Hannover is presented as an example.

1.6.1 Hydrology: Groundwater Variations in Hannover

Figure 1.18 shows the time series of absolute gravity determinations in Hannover (point 103) observed with the two Hannover instruments. The station is located on glacial sediments with a thickness of about 500 m (sand, clay and marl of low consolidation). The free-fall experiments are severely affected by natural (wind forces on the adjacent buildings) and man-made (machines, traffic) seismic noise. The history of the Hannover measurements reveals a linear gravity decrease of about 25 μGal over a period of 21 years, whereas in Clausthal (Fig. 1.6) no significant secular gravity variation can be found. An explanation for the phenomenon in Hannover is not yet available and requires discussions with other experts, e.g. from hydrosphere research. Figure 1.19 illustrates the scatter in the time histories which is not

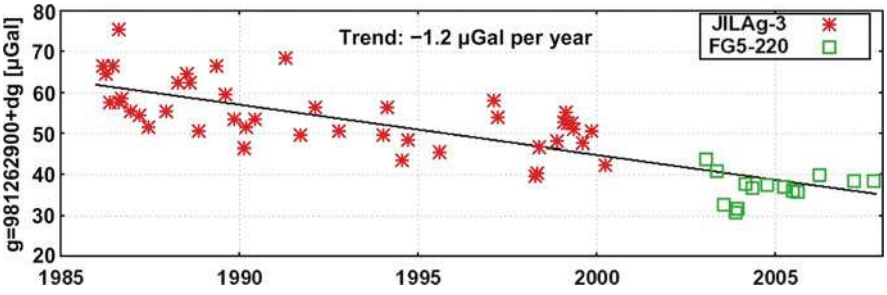


Fig. 1.18 Absolute gravity determinations with JILAg-3 (offset corrected) and FG5-220 at stations Hannover (HAN103, trend -1.2 ± 0.1 μGal/year)

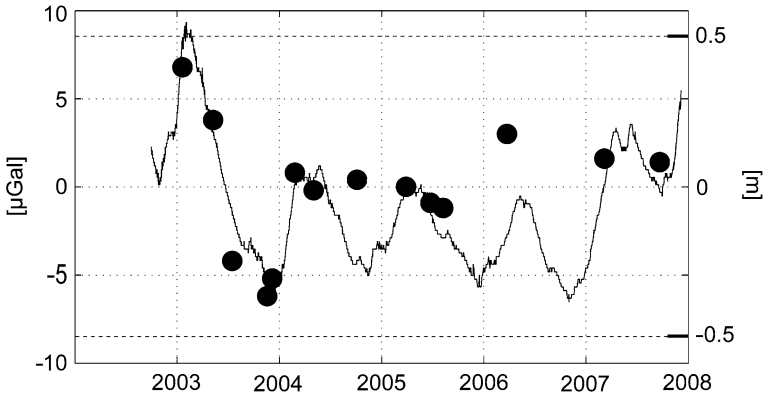


Fig. 1.19 Groundwater table at the gravimetry laboratory in Hannover and absolute gravity determinations with FG5-220 since 2003. The transfer function from gravity to groundwater change, with the linear coefficient 17 µGal/m (correlation 80%), has been applied to the absolute gravity determinations to convert the *g*-values to groundwater readings

only caused by measurement uncertainties but also by real gravity variations. For example, from February to December 2003 the groundwater table at the gravimetry laboratory in Hannover fell about 70 cm due to the very dry season in northern Germany. This was accompanied by a gravity decrease of 13 µGal. Checking the groundwater readings for the period 1986 to the present, a declining trend over the years is not visible. But these readings from the groundwater gauge consider only the upper aquifer of the subsurface hydrology around the gravimetry laboratory and not the deeper aquifers. Thus, it cannot be excluded that the long-term trend in the gravity series might be caused by a change in the subsurface water content.

1.6.2 Tectonics: Isostatic Land Uplift in Fennoscandia

In the Fennoscandian land uplift area, the Earth's crust has been rising continuously since the last glacial maximum in response to the deloading of the ice. This process is an isostatic adjustment of the Earth's elastic lithosphere and underlying viscous mantle. For a general overview Wolf (1993) gives a historical review about the changing role of the lithosphere in models of glacial isostasy.

The Fennoscandian rebound area is dominated by the Precambrian basement rocks of the Baltic Shield, which is part of the ancient East European Craton and comprises south Norway, Sweden, Finland, the Kola Peninsula and Russian Karelia. The region is surrounded by a flexural bulge, covering northern Germany and northern Poland, the Netherlands and some other surrounding regions. The bulge area was once rising due to the Fennoscandian ice load and, after the melting, sinking with a much smaller absolute value than the uplift rate in the centre of Fennoscandia. Denmark is part of the transition zone from the uplift to the subsidence area. The maximum spatial extension of the uplift area is about 2,000 km in northeast–southwest direction; see Fig. 1.20 for the approximate shape (after Ågren

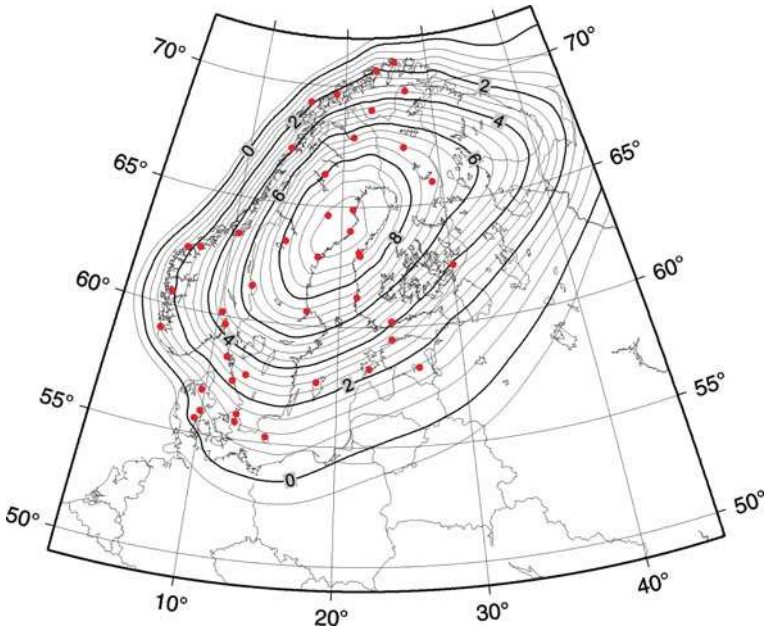


Fig. 1.20 Map of the postglacial uplift of Fennoscandia in mm/year after Ågren and Svensson (2007) derived from model NKG2005LU, courtesy of Ågren. The *dots* indicate the positions of gravity stations of the Nordic Absolute Gravity Project

and Svensson 2007). Presently, the central area around the northern part of the Gulf of Bothnia is undergoing an uplift at a rate of about 1 cm/year.

The Trans-European Suture Zone (TESZ) is a main tectonic boundary in Europe, separating the East European Craton from the Phanerozoic terrains in the west and southwest (Palaeozoic western Europe and Meso-Europe). The Sorgenfrei-Tornquist Zone is part of the TESZ and crosses Denmark north of Copenhagen in the immediate vicinity of the absolute gravity station Helsingør; see Fig. 1.10 for the station names and locations. Among other stations, the absolute gravity stations Copenhagen/Vestvolden, Helsingør, Onsala and Borås belong to the Nordic Geodetic Observing System (NGOS) and constitute the central part of a north–south profile crossing perpendicularly the graben system of the suture zone between the Baltic Shield and the younger Palaeo-Europe.

Four east–west profiles across the Fennoscandian postglacial rebound area have been utilised by relative gravimetry and levelling. They follow approximately the latitudes 65°N (observed 1975–2000), 63°N (1966–2003), 61°N (1976–1983) and 56°N (1977–2003); see Ekman and Mäkinen (1996) or Mäkinen et al. (2004). The east–west directions were chosen to ensure only small gravity differences between the relative gravimetry points (less than 1 mGal). This requirement avoids errors from uncertainties of the gravimeter calibrations. With the availability of transportable absolute gravimetry in Central Europe, the 56° profile (Denmark–Sweden) was supported with JILAg-3 (in 1986) and FG5-220 (2003, 2005) measurements.

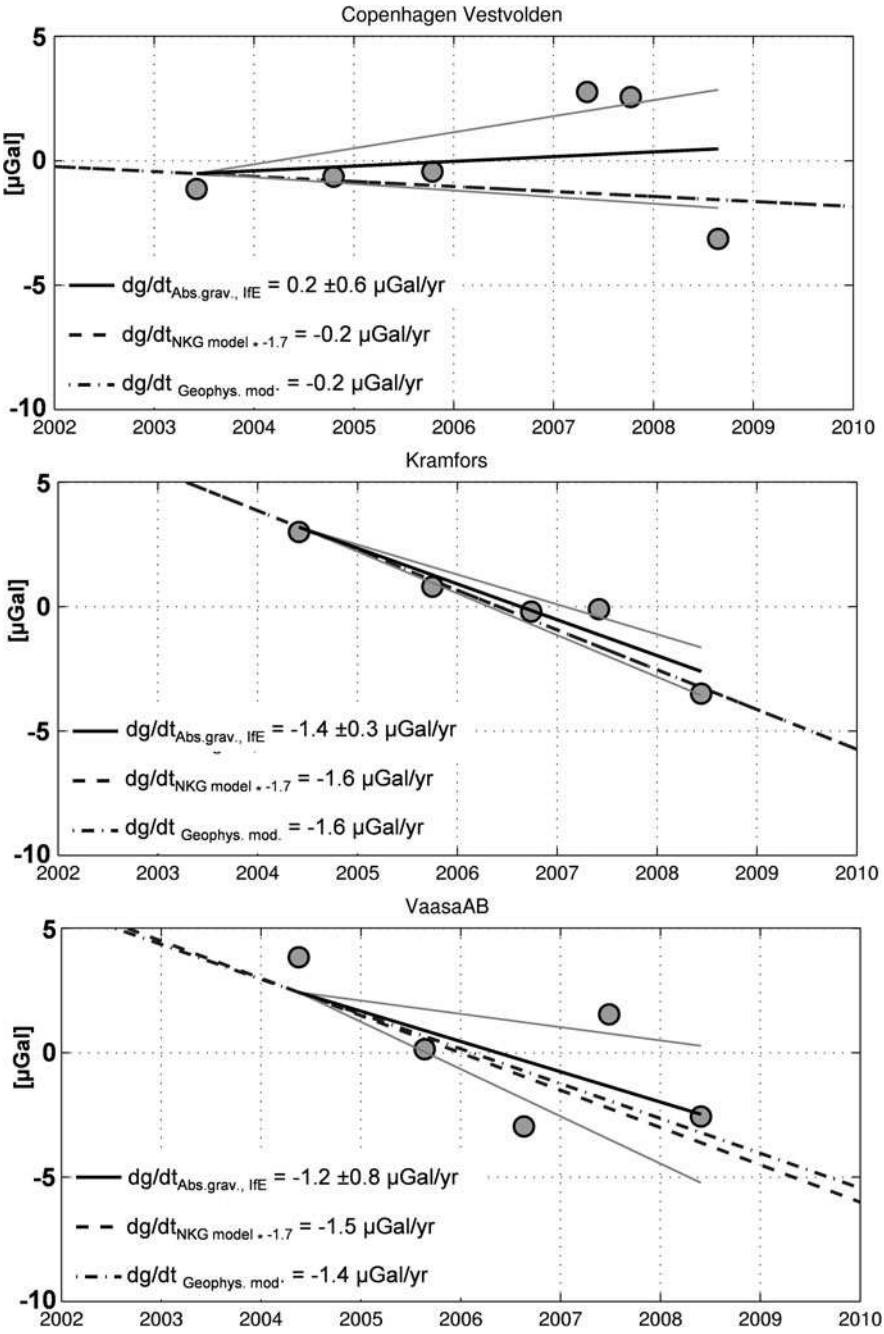


Fig. 1.21 Linear gravity changes for three stations in Fennoscandia derived from absolute gravity measurements of IfE and compared with the trends from the NKG2005LU model (Fig. 1.20, conversion factor −1.7 µGal/cm applied) and from the model predictions provided by Klemann (2004). The *grey lines* beside the IfE trends indicate the standard deviation of the corresponding trend line

Repeated observations with the FG5-220 from IfE were performed at 11 stations in Fennoscandia nearly every year from 2003 to 2008 (Gitlein et al. 2008). From these results, linear gravity changes were calculated for each station (Gitlein 2009). The trends for three stations are presented exemplarily in Fig. 1.21. A decrease in gravity due to land uplift is evident at almost all stations. The largest gravity changes were found around the uplift centre (e.g. Kramfors). In Copenhagen, close to the zero uplift line in Fig. 1.20, the obtained gravity rate is nearly zero. Overall, the regional rebound signal is clearly visible, but still seems to be disturbed by environmental mass variations, e.g. at station VaasaAB. From the experiences over the last 5 years, the hydrological changes are considered as a main contributor, which is also indicated by the water level observations of the reservoirs and wells close to some of the absolute stations.

In Fig. 1.21 the observational trends are compared with the results derived from the NKG2005LU model (Ågren and Svensson 2007), which is mainly based on levelling and GPS results, and with predicted rates of a glacial rebound model provided by Klemann (2004). The computations were based on solution algorithms developed by Martinec (2000) and Hagedoorn et al. (2007) and use a global ice model with SCAN-II (Lambeck et al. 1998) for Fennoscandia. Overall, the trends observed by absolute gravimetry since 2003 or 2004 are in good agreement with the other results. The obtained standard deviations seem to be realistic estimates for the accuracy of the deduced secular gravity changes. The disturbances caused by unaccounted hydrological effects are cancelled out in the trends to some extent due to the annual gravity measurements. Thus, absolute gravimetry has shown its capability to observe the Fennoscandian land uplift within the rather short time span of 4–5 years.

References

Agnew, D.C. (1997) NLOADF: a program for computing ocean-tide loading. J. Geophys. Res., 102(B3), 5109–5110

Ågren, J. and Svensson, R. (2007) Postglacial land uplift model and system definition for the new Swedish height system RH 2000. Reports in Geodesy and Geographical Information Systems, Lantmateriet, Gävle, 123 pp

Bilker-Koivula, M., Mäkinen, J., Timmen, L., Gitlein, O., Klopping, F. and Falk, R. (2008) Repeated absolute gravity measurements in Finland. In: Peshekhonov, V.G. (ed) Terrestrial Gravimetry: Static and Mobile Measurements (TG-SMM2007). Proceedings of International Symposium, Elektropribor, St. Petersburg, pp. 147–151

BIPM (2006) Bureau international des poids et measures, Le Système international d'unités (SI) – The International System of Units (SI). 8e édition, 2006. Édité par le BIPM, Pavillon de Breteuil, F-92312 Sèvres Cedex, France

Boedecker, G. and Flury, J. (1995) International Absolute Gravity Basestation Network IAGBN, Catalogue of stations and observations. Report of the IAG International Gravity Commission, IGC-Working Group 2, "World Gravity Standards", available at the Bureau Gravimetrique International, Toulouse

Boedecker, G. and Fritzer, Th. (1986) International Absolute Gravity Basestation Network. Veröff. Bayer. Komm. für die Internat. Erdmessung der Bayer. Akad. d. Wissensch., Astron.-Geod. Arb. 47, München

Boulanger, Yu., Faller, J., Groten, E., Arnautov, G., Becker, M., Bernard, B., Cannizzo, L., Cerutti, G., Courtier, N., Feng Youg-Yuan, Friederich, J., Guo You-Guang, Hanada, H., Huang Da-Lun, Kalish, E., Klopping, F., Li De-Xi, Liard, J., Mäkinen, J., Marson, I., Ooe, M., Peter, G., Röder, R., Ruess, D., Sakuma, A., Schnüll, M., Stus, F., Scheglov, S., Tarasuk, W., Timmen, L., Torge, W., Tsubokawa, T., Tsuruta, S., Vänskä, A. and Zhang Guang-Yuan (1991) Results of the 3rd International Comparison of Absolute Gravimeters in Sèvres 1989. Bur. Grav. Int., Bull. d'Inf., 68, 24–44, Toulouse

Carter, W.E., Aubrey, D.G., Baker, T., Boucher, C., LeProvost, C., Pugh, D., Peltier, W.R., Zumberge, M., Rapp, R.H., Schutz, R.E., Emery, K.O. and Enfield, D.B. (1989) Geodetic fixing of tide gauge bench marks. Woods Hole Oceanographic Institution Report WHOI-89-31/CRC-89-5, Woods Hole

Carter, W.E., Peter, G., Sasagawa, G.S., Klopping, F.J., Berstis, K.A., Hilt, R.L., Nelson, P., Christy, G.L., Niebauer, T.M., Hollander, W., Seeger, H., Richter, B., Wilmes, H. and Lothammer, A. (1994) New gravity meter improves measurements. EOS, Trans. Am. Geophys. Union, 75(08), 90–92

Cartwright, D.E. and Edden, A.C. (1973) Corrected tables of tidal harmonics. Geophys.J. R. Astr. Soc., 33, 253–264

Cartwright, D.E. and Tayler, R.J. (1971) New computations of the tide-generating potential. Geophys. J. R. Astr. Soc., 23, 45–74

Cook, A.H. (1965) The absolute determination of the acceleration due to gravity. Metrologia, 1, 84–114

Dehant, V. (1987) Tidal parameters for an inelastic Earth. Phys. Earth Planet. Inter., 49, 97–116

Doodson, A.T. (1921) The harmonic development of the tide-generating potential. Proc. R. Soc. (London), Series A 100, 306–328. Reprint in Int. Hydrographic Revue 31, Monaco 1954

Dziewonski, A.M. and Anderson, D.L. (1981) Preliminary reference earth model (PREM). Phys. Earth Planet. Int., 25, 297–356

Ekman, M. and Mäkinen, J. (1996) Recent postglacial rebound, gravity change and mantle flow in Fennoscandia. Geophys. J. Int., 126, 229–234

Everaerts, M., Lambot, Ph., Van Hoolst, T., van Ruymbeke, M. and Ducarme, B. (2002) First order gravity network of Belgium. Bur. Grav. Int., Bull. d'Inf., 90, 27–42, Toulouse

Falk, R. (1995) Erste Erfahrungen mit dem Automatischen Gravimeter Scintrex CG-3 M Autograv. Zeitschrift für Vermessungswesen, Verlag Wittwer, Stuttgart, Heft 1, 26–34

Faller, J.E., Guo, Y.G., Geschwind, J., Niebauer, T.M., Rinker, R.L. and Xue, J. (1983) The JILA portable absolute gravity apparatus. Bur. Grav. Int., Bull. d'Inf., 53, 87–97, Toulouse

Francis, O. and van Dam, T. (2006) Analysis of results of the International Comparison of Absolute Gravimeters in Walferdange (Luxembourg) of November 2003. Cahiers du Centre Européen de Géodynamique et de Séismologie, 26, 1–23, Luxembourg

Francis, O., van Dam, T., Germak, A., Amalvict, M., Bayer, R., Bilker-Koivula, M., Calvo, M., D'Agostino, G.-C., Dell'Acqua, T., Engfeldt, A., Faccia, R., Falk, R., Gitlein, O., Fernandez, Gjevestad, J., Hinderer, J., Jones, Kostelecky, J., Le Moigne, N., Luck, B., Mäkinen, J., McLaughlin, D., Olszak, T., Olsson, P., Pachuta, A., Palinkas, V., Pettersen, B., Pujol, R., Prutkin, I., Quagliotti, D., Reudink, R., Rothleitner, C., Ruess, D., Shen, C., Smith, V., Svitlov, S., Timmen, L., Ulrich, C., Van Camp, M., Walo, J., Wang, L., Wilmes, H. and Xing, L. (2010) Results of the European Comparison of Absolute Gravimeters in Walferdange (Luxembourg) of November 2007. Proceedings of the IAG Symposium on Gravity, Geoid and Earth Observations 2008, Chania, Crete, Greece, June 2008, Springer (accepted for publication)

Gitlein, O. (2009) Absolutgravimetrische Bestimmung der Fennoskandischen Landhebung mit dem FG5-220. Dissertation, Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover (im Review-Prozess)

Gitlein, O. and Timmen, L. (2006) Atmospheric mass flow reduction for terrestrial absolute gravimetry in the Fennoscandian land uplift network. In: Tregoning, P. and Rizos, C. (eds) Dynamic Planet. IAG Symposium 130, Springer, Berlin, pp. 461–466

Gitlein, O., Timmen, L., Müller, J., Denker, H., Mäkinen, J., Bilker-Koivula, M., Pettersen, B.R., Lysaker, D.I., Svendsen, J.G.G., Wilmes, H., Falk, R., Reinhold, A., Hoppe, W., Scherneck, H.-G., Engen, B., Omang, O.C.D., Engfeldt, A., Lilje, M., Strykowski, G. and Forsberg, R. (2008) Observing absolute gravity acceleration in the Fennoscandian land uplift area. In: Peshekhonov, V.G. (ed) Terrestrial Gravimetry: Static and Mobile Measurements (TG-SMM2007). Proceedings of the International Symposium. Elektropribor, St. Petersburg, pp. 175–180

Hagedoorn, J.M., Wolf, D. and Martinec, Z. (2007) An estimate of global mean sea-level inferred from tide-gauge measurements using glacial-isostatic models consistent with the relative sea-level record. Pure Appl. Geophys., 164, 791–818, Birkhäuser Verlag, Basel

Hartmann, T. and Wenzel, H.-G. (1995) The HW95 tidal potential catalogue. Geophys. Res. Lett., 22, 3553–3556

Hugill, A.L. (1988) The new Scintrex CG-3 Automated Gravity Meter: description and test results. Paper presented at the ASEG/SEG Conference, February 1988. Adelaide

IGC (1988) International Absolute Gravity Basestation Network (IAGBN), Absolute Gravity Observations, Data Processing Standards & Station Documentation (Int. Grav. Com. -WGII: World Gravity Standards). Bur. Grav. Int., Bull. d'Inf., 63, 51–57, Toulouse

Ilk, K.H., Flury, J., Rummel, R., Schwintzer, P., Bosch, W., Haas, C., Schröter, J., Stammer, D., Zahel, W., Miller, H., Dietrich, R., Huybrechts, P., Schmeling, H., Wolf, D., Götze, H.J., Riegger, J., Bardossy, A., Güntner, A. and Gruber, Th. (2005) Mass transport and mass distribution in the Earth system – Contribution of the new generation of satellite gravity and altimetry missions to geosciences. GOCE Projektbüro, TU München und GFZ Potsdam

Jentzsch, G. (1997) Earth tides and ocean tidal loading. In: Wilhelm, H., Zürn, W. and Wenzel, H.-G. (eds) Tidal Phenomena. Lecture Notes in Earth Sciences, Vol. 66. Springer, Heidelberg, pp. 145–171

Jentzsch, G. (2008) The automated Burris Gravity Meter – a new instrument using an old principle. In: Peshekhonov, V.G. (ed) Terrestrial Gravimetry: Static and Mobile Measurements (TG-SMM2007). Proceedings of the International Symposium. Elektropribor, St. Petersburg, pp. 21–28

Jousset, M., van Ruymbeke, M., Bonvalot, S. and Diament, M. (1995) Performance of two Scintrex CG3M instruments at the Fourth International Comparison of Absolute Gravimeters. Metrologia, Sèvres, 32, 231–244

Kanngieser, E., Kummer, K., Torge, W. and Wenzel, H.-G. (1983) Das Gravimeter Eichsystem Hannover. Wiss. Arb. d. Fachrichtung Verm.wesen der Univ. Hannover, Nr. 120

Klemann, V. (2004) Linear gravity variations for the IfE absolute gravity stations in Fennoscandia predicted by geophysical GIA modelling. Personal communication, Deutsches GeoForschungsZentrum (GFZ), Potsdam

Krieg, L.A. (1981) Mathematical modelling of the behaviour of the LaCoste and Romberg "G" gravity meter for use in gravity network adjustments and data analyses. Reports of the Department of Civil and Environmental Engineering and Geodetic Science, Geodetic Science and Surveying, The Ohio State University, Columbus, 321

Lambeck, K., Smither, C. and Ekman, M. (1998) Tests of glacial rebound models for Fennoscandinavia based on instrumented sea- and lake-level records. Geophys. J. Int., 135, 375–387

Lambert, A., Courtier, N., Sasagawa, G.S., Klopping, F., Winester, D., James, T.S. and Liard, J.O. (2001) New constraints on Laurentide Postglacial rebound from absolute gravity measurements. Geophys. Res. Lett., 28, 2109–2211

Lambert, A., James, T.S., Liard, J.O. and Coutier, N. (1996) The role and capability of absolute gravity measurements in determining the temporal variations in the earth's gravity field. IAG Symposium 116. Springer, Berlin, pp. 20–29

Liard, J. and Gagnon, C. (2002) The new A-10 absolute gravimeter at the 2001 International Comparison of Absolute Gravimeters. Metrologia, 39, 477–483

Mäkinen, J., Amalvict, M., Shibuya, K. and Fukuda, Y. (2007) Absolute gravimetry in Antarctica: status and prospects. J Geodynamics, 43, 339–357

Mäkinen, J., Engfeldt, A., Harrson, B.G., Ruotsalainen H., Strykowski, G., Oja, T. and Wolf, D. (2004) The Fennoscandian Land Uplift Gravity Lines 1966–2003. Proceedings (CD publ.) of the IAG Symposium Gravity, Geoid and Space Missions (GGSM2004), Porto, 2004

Mäkinen, J. and Tattari, S. (1988) Soil moisture and groundwater: two sources of gravity variations. Bur. Grav. Int., Bull. d'Inf., 62, 103–110, Toulouse

Marson, I., Faller, J.E., Cerutti, G., De Maria, P., Chartier, J.-M., Robertsson, L., Vitushkin, L., Friederich, J., Krauterbluth, K., Stizza, D., Liard, J., Gagnon, C., Lothhammer, A., Wilmes, H., Mäkinen, J., Murakami, M., Rehren, F., Schnüll, M., Ruess, D. and Sasagawa, G.S. (1995) Fourth international comparison of absolute gravimeters. Metrologia, 32, 137–144

Martinec, Z. (2000) Spectral-finite element approach to three-dimensional viscoelastic relaxation in a spherical earth. Geophys. J Int, 142, 117–141

Micro-g Solutions Inc. (1999) Operator's manual, FG5 absolute gravimeter. Micro-g Solutions Inc, Erie

Müller, J., Neumann-Redlin, M., Jarecki, F., Denker, H. and Gitlein, O. (2006) Gravity changes in northern Europe as observed by GRACE. In: Tregoning, P. and Rizos, C. (eds) Dynamic Planet. IAG Symposium 130. Springer, Berlin, pp. 523–257

Nagornyi, V.D. (1995) A new approach to absolute gravimeter analysis. Metrologia, 32, 201–208

Niebauer, T.M. (1987) New absolute gravity instruments for physics and geophysics. Ph.D. thesis, University of Colorado, Department of Physics, Boulder

Niebauer, T.M., Klopping, F.G., Bilson, R. and Brown, J.M. (1999) The new A10 absolute field gravimeter. EOS Trans. AGU, 80(17), Spring Meet. Suppl. S85

Niebauer, T.M., Sasagava, G.S., Faller, J.E., Hilt, R. and Klopping, F. (1995) A new generation of absolute gravimeters. Metrologia, 32, 159–180

Pearlman, M., Altamini, Z., Beck, N., Forsberg, R., Gurtner, W., Kenyon, S., Behrend, D., Lemoine, F.G., Ma, C., Noll, C.E., Pavlis, E.C., Malkin, Z., Moore, A.W., Webb, F.H., Neilan, R.E., Ries, J.C., Rothacher, M. and Willis, P. (2006) GGOS working group on ground networks and communications. In: Tregoning, P. and Rizos, C. (eds) Dynamic Planet. IAG Symposium, 130, Springer, Berlin, pp. 719–726

Rapp, R.H. (1983) Tidal gravity computations based on recommendations of the Standard Earth Tide Committee. Bull. d'Inf., Marées Terrestres, 89, 5814–5819, Bruxelles

Rehren, F. (1997) Relative gravity measurements with a Scintrex CG-3 M in the Gravimeter Calibration Systems Hannover and Hornisgrinde. Bur. Grav. Int., Bull. d'Inf., 81, 23–29, Toulouse

Reigber, C. and Feissel, M. (1997) IERS missions, present and future. Report on the 1996 IERS Workshop. IERS Technical Note 22, Paris

Robertsson, L., Francis, O., vanDam, T.M., Faller, J., Ruess, D., Delinte, J.-M., Vitushkin, L., Liard, J., Gagnon, C., Guo You Guang, Huang Da Lun, Fang Yong Yuan, Xu Jin Yi, Jeffries, G., Hopewell, H., Edge, R., Robinson, I., Kibble, B., Mäkinen, J., Hinderer, J., Amalvict, M., Luck, B., Wilmes, H., Rehren, F., Schmidt, K., Schnüll, M., Cerutti, G., Germak, A., Zabek, Z., Pachuta, A., Arnautov, G., Kalish, E., Stus, Y., Stizza, D., Friederich, J., Chartier, J.-M. and Marson, I. (2001) Results from the Fifth International Comparison of Absolute Gravimeters, ICAG97. Metrologia, 38, 71–78

Röder, R.H., Schnüll, M. and Wenzel, H.-G. (1985) Gravimetry with an electrostatic feedback system. Bur. Grav. Int., Bull. d'Inf., 57, 72–81, Toulouse

Röder, R.H., Schnüll, M. and Wenzel, H.-G. (1988) SRW feedback for LaCoste-Romberg gravimeters with extended range. Bur. Grav. Int., Bull. d'Inf., 62, 46–50, Toulouse

Romagnoli, C., Zerbini, S., Lago, L., Richter, B., Simon, D., Domenichini, F., Elmi, C. and Ghirotti, M. (2003) Influence of soil consolidation and thermal expansion effects on height and gravity variations. J Geodynamics, 35, 521–539

Schmerge, D. and Francis, O. (2006) Set standard deviations, repeatability and offset of absolute gravimeter A10-008. Metrologia, 43, 414–418

Schnüll, M., Röder, R.H. and Wenzel, H.-G. (1984) An improved electronic feedback for LaCoste-Romberg gravity meters. Bur. Grav. Int., Bull. d'Inf., 55, 27–36, Toulouse

Schuh, H., Dill, R., Greiner-Mai, H., Kutterer, H., Müller, J., Nothnagel, A., Richter, B., Rothacher, M., Schreiber, U. and Soffel, M. (2003) Erdrotation und globale dynamische Prozesse. Mitteilungen des Bundesamtes für Kartographie und Geodäsie 32, 118 Seiten, Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt a.M.

Schwiderski, E. (1980) Ocean tides, part I: global ocean tidal equations; part II: a hydrodynamical interpolation model. Marine Geodesy, 3, 161–255

Scintrex (1998) CG-3/3 M Gravity Meter, User's Guide, Concord

Tamura, Y. (1987) A harmonic development of the tide generating potential. Bull. d'Inf., Marées Terrestres, 99, 6813–6855, Bruxelles

Tapley, B., Bettadpur, S., Ries, J., Thompson, P. and Watkins, M. (2004) GRACE measurements of mass variability in the Earth System. Science, 305, 503–505

Timmen, L. (1994) Untersuchungen zur Modellbildung bei der Auswertung absoluter Schweremessungen. Wiss. Arb. d. Fachrichtung Verm.wesen der Univ. Hannover, 204, Dissertation am Institut für Erdmessung, Universität Hannover

Timmen, L. (1996) Absolutgravimetrie – Aufgaben und Anwendungen. Zeitschrift für Vermessungswesen (ZfV) 121, 286–295, Wittwer, Stuttgart

Timmen, L. (2003) Precise definition of the effective measurement height of free-fall absolute gravimeters. Metrologia, 40, 62–65

Timmen, L., Flury, J., Peters, T. and Gitlein, O. (2006a) A new absolute gravity base in the German Alps. In: Hvoždara, M. and Kohúh I. (eds) Contributions to Geophysics and Geodesy, 2nd Workshop on International Gravity Field Research (special issue) 36, 7–20, Slovak Academy of Sciences, Bratislava

Timmen, L., Gitlein, O., Müller, J., Denker, H., Mäkinen, J., Bilker, M., Pettersen, B.R., Lysaker, D.I., Omang, O.C.D., Svendsen, J.G.G., Wilmes, H., Falk, R., Reinhold, A., Hoppe, W., Scherneck, H.-G., Engen, B., Harsson, B.G., Engfeldt, A., Lilje, M., Strykowski, G. and Forsberg, R. (2006b) Observing fennoscandian gravity change by absolute gravimetry. In: Sansò, F. and Gil A.J. (eds) Geodetic Deformation Monitoring: From Geophysical to Engineering Roles. IAG Symposium 131, Springer, Berlin, Heidelberg, pp. 193–199

Timmen, L. and Wenzel, H.-G. (1994) Improved gravimetric Earth tide parameters for station Hannover. Bull. d'Inf., Marées Terrestres, 119, 8834–8846, Bruxelles

Timmen, L. and Wenzel, H.-G. (1995) Worldwide synthetic gravity tide parameters. In: Sünkel, H. and Marson, I. (eds) Gravity and Geoid. Proceedings of IAG Symposium, 113, Springer, Berlin, Heidelberg, pp. 92–101

Torge, W. (1989) Gravimetry. de-Gruyter, Berlin-New York

Torge, W. (1990) Absolute gravimetry as an operational tool for geodynamic research. In: Brunner, F.K. and Rizos, C. (eds) Proceedings of the Ron Mather Symposium on Four-Dimensional Geodesy (Sydney, 1989). Lecture Notes in Earth Sciences, Vol. 29. Springer, Berlin, pp. 15–28

Torge, W. (1991) The present state of absolute gravimetry. Cahiers du Centre Européen de Géodynamique et de Séismologie, 3, 9–22, Luxembourg

Torge, W. (1993) Gravimetry and tectonics. Publications of the Finnish Geodetic Institute (Geodesy and Geophysics) 115, 131–172, Helsinki

Torge, W. (1998a) The changing role of gravity reference networks. In: Forsberg, R., Feissel, M. and Dietrich, R. (eds) Geodesy on the Move. IAG Symposium, 119, Springer, Berlin, Heidelberg, New York, etc., pp. 1–10

Torge, W. (1998b) 100 Jahre Schwerereferenznetze – klassische und moderne Konzeption. Zeitschrift für Vermessungswesen (ZfV), 123, 355–363, Wittwer, Stuttgart

Torge, W., Falk, R., Franke, A., Reinhart, E., Richter, B., Sommer, M. and Wilmes, H. (1999) Das Deutsche Schweregrundnetz 1994 (DSGN94). DGK, B 309, München

Torge, W., Röder, R.H., Schnüll, M., Wenzel, H.-G. and Faller, J.E. (1987) First results with the transportable absolute gravity meter JILAg-3. Bull. Géod., 61(2), 161–176, Springer, Berlin – Heidelberg

Wahr, J.M. (1981) Body tides on an elliptical, rotating, elastic and oceanless Earth. Geophys. J. Royal Astron. Soc., 64, 677–703, Oxford

Wahr, J. (1985) Deformation induced by polar motion. J. Geophys. Res., 92(B2), 1281–1286
Wahr, J. and Velicogna, I. (2003) What might GRACE contribute to studies of post glacial rebound. Space Sci. Rev., 108, 319–330
Wang, R. (1997) Tidal response of the solid Earth. In: Wilhelm, H., Zürn, W. and Wenzel, H.-G. (eds) Tidal Phenomena. Lecture Notes in Earth Sciences, Vol. 66, Springer, Heidelberg, pp. 27–57
Wenzel, H.-G. (1993) Programmsystem GRAV, programme manual, Geodetical Institute University of Karlsruhe, Karlsruhe
Wenzel, H.-G. (1997) Tide-generating potential for the Earth. In: Wilhelm, H., Zürn, W. and Wenzel, H.-G. (eds) Tidal Phenomena. Lecture Notes in Earth Sciences, Vol. 66. Springer, Heidelberg, pp. 9–26
Williams, S.D.P., Baker. T.F. and Jeffries, G. (2001) Absolute gravity measurements at UK tide gauges. Geophys. Res. Lett., 28, 2317–2329.
Wilmes, H. and Falk, R. (2006) Bad Homburg – a regional comparison site for absolute gravity meters. Cahiers du Centre Européen de Géodynamique et de Séismologie, 26, 29–30, Luxembourg
Wolf, D. (1993) The changing role of the lithosphere in models of glacial isostasy: a historical review. Global Planetary Change, 8, 95–106, Elsevier Science Publishers B.V., Amsterdam
Xu, J., Zhu, S., Lui, X., Torge, W., Röder, H., Schnüll, M. and Wenzel, H.-G. (1988) Vertical Gravimeter Calibration Line Wuhan/China. Bur. Grav. Int., Bull. d'Inf., 62, 119–125, Toulouse
ZLS Corporation (2007) User guide: Automated Burris Gravity Meter™ and UltraGrav ™ Control System, 67pp
Zürn, W. and Wilhelm, H. (1984) Tides of the earth. Landolt-Börnstein, New Series V, 2, 259–299, Springer, Heidelberg-New-York

Chapter 2
Adaptively Robust Kalman Filters with Applications in Navigation

Yuanxi Yang

Contents

Y. Yang (✉)
Xian Research Institute of Surveying and Mapping, Xian 710054, China
e-mail: yuanxi@pub.xaonline.com

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_2,

A new adaptively robust Kalman filtering was developed in 2001. The main achievements of the adaptively robust filter are summarized from the published papers in recent years. These include the establishment of the principle of the adaptively robust filter, the derivation of the corresponding state parameter estimator, the developments of four adaptive factors for balancing the contribution of kinematic model information and measurements, which include three-segment function, two-segment function, exponential function and zero and one function for state component adaptation, and the establishment of four kinds of learning statistics for judging the kinematic model errors, which include state discrepancy statistic, predicted residual statistic, variance component ratio statistic and velocity discrepancy statistic. The relations of the adaptively robust filter with standard Kalman filter, robust filter and some other adaptive Kalman filters as well as some related adjustment methods are depicted by a figure. Other developments of the adaptively robust filter are also presented.

2.1 Introduction

Applications of the Kalman filter in dynamic or kinematic positioning have sometimes encountered difficulties which have been referred to as divergences. These divergences can often be traced to three factors: (1) insufficient accuracy in modelling the dynamics or kinematics (functional model errors of the state equations); (2) insufficient accuracy in modelling the observations (functional model errors of observation equations); and (3) insufficient accuracy in modelling the distributions or the priori covariance matrices of the measurements and the updated parameters (stochastic model errors) (Yang et al. 2001a).

The current basic procedure for the quality control of Kalman filter consists of

- Functional model compensation for model errors by introducing uncertain parameters into the state and/or the observation equations. Any model error term can be introduced into the models arbitrarily. One could then augment the state (Jazwinski 1970, p. 308). A similar approach is developed by Schaffrin (1991, pp. 32–34). He partitions the state vector into h groups, each being affected by a common scale error. Then an $h \times 1$ vector of scale parameters is introduced into the models. This kind of approach may, of course, lead to a high-dimensional state vector which, in turn, greatly increases the filter computational load (Jazwinski 1970, p. 305).
- Stochastic model compensation by introducing a variance–covariance matrix of the model errors. In taking this approach to prevent divergence, one has to determine what covariance matrix to add. A reasonable covariance matrix may compensate for the model errors. An ineffective covariance matrix, however, adds the model divergence. For instance, when the model is accurate in some dynamic or kinematic periods, an unsuitable increasing of the covariance matrix of model

error will degrade the state estimators. An effective covariance matrix for model errors can only be determined by trial and error.

- DIA procedure – detection, identification and adaptation (Teunissen 1990). It uses a recursive testing procedure to eliminate outliers. In the detection step one looks for unspecified model errors. In the identification step one tries to find the cause of the model error and its most likely starting time. After a model error has been detected and identified, the bias in the state estimate caused by the model error has to be eliminated as well. This model recovery from errors is called adaptation (Salzmanm 1995). The identification of the model, however, is quite difficult, especially when the measurements are not accurate enough to detect the unspecified model errors.
- Sequential least squares procedure. A quite different procedure that has been frequently used for kinematic positioning does not use the dynamic model information at all but determines discrete positions at the measurement epochs (Cannon et al. 1986). In this case, no assumption on dynamic model is made; only the measurements at discrete epoch are employed to estimate the state parameters. The model error, therefore, does not affect the estimates of new state parameters. Usually, this method is presented as a sequential least squares algorithm (Schwarz et al. 1989). The current limitation of this approach is that it wastes the good information of the state model when the model accurately describes the dynamic process in cases.
- Sage adaptive Kalman filtering. This kind of adaptive filter evaluates the variance–covariance matrices of the kinematic model error vector and the measurement error vector by windowing method (Sage and Husa 1969). In the applications, an innovation-based adaptive Kalman filtering for an integrated INS/GPS is developed by Mohamed and Schwarz (Mohamed and Schwarz 1999; Wang et al. 2000). The problem is that the algorithm needs to collect the residuals of the measurements or the update series to calculate the underlined variance–covariance matrices; thus it requires that the measurement dimensions and types at all epochs be the same.
- Fading Kalman filtering. In order to control the influences of prior state errors or kinematic model errors on the present estimated state parameters, the fading filters, using the fading factors to restrict the memory length of Kalman filter and to make the most use of present measurements, were developed in the field of statistics as early as the 1960s and 1970s (Fagin 1964; Sorenson and Sacks 1971). We have analysed the basic properties of the fading filter, the abilities in controlling the influences of the kinematic model errors on the state parameter estimates and the problems possibly existing in the practical applications (Yang and Gao 2006c).
- Robust filter based on min–max robust theory. The deviation of observation error distribution from the Gaussian one may also seriously degrade the performance of the Kalman filtering. Thus, there appears to be considerable motivation for considering filters which are robustised to perform fairly well in non-Gaussian environments. Facing this problem, Masreliez and Martin (1977) applied the influence function of min–max robust theory to replace the score function of the standard Kalman filter. The basic disadvantages associated with this

kind of robust filter are that the estimator requires the unknown contaminating distribution to be symmetric and it cannot work as well as the standard Kalman filter in Gaussian noise.

- Robust filter based on M estimation theory (Huber 1964) and Bayesian statistics. To resist the bad influences of both state model errors and measurement outliers, a robust M–M filter is developed (Yang 1991, 1993, 1997, 1999; Zhou et al. 1997, p. 299) by which the measurement outliers are controlled by robust equivalent weights of the measurements, and the model errors are resisted by the equivalent weights of the update parameters according to the divergence of the predicted parameters and the estimated ones. Furthermore, a robust filter for rank-deficient observation models was developed by Koch and Yang (1998) by Bayesian statistics and by applying the robust M-estimate.

Different from Sage–Husa adaptive filtering (see Deng 2003, pp. 162–173; Mohamed and Schwarz 1999; Wang and Kubik 1993; Wang et al. 2000) and limited memory filter (Panozzo et al. 2004) as well as the mentioned adaptive filters, a new adaptively robust filter was developed by combining the adaptive Kalman filter and robust estimation (Yang et al. 2001a), which applies a robust estimation principle for measurement vector to resist its outlier effects and introduces an adaptive factor for the model predicted state vector to control its outlying disturbance influences.

After adaptively robust filtering was developed, four learning statistics and four adaptive factors have been set up based on experiences and have been proved effective in practical applications. An accompanying adaptive factor with a three-segment descending function and a learning statistic constructed by using the discrepancy between the predicted state from the kinematic model and the state estimated from the measurements have been established. Three other kinds of adaptive factors have been developed, which are a two-segment descending function (Yang et al. 2001b), an exponential function (Yang and Gao 2005) and a zero and one function for state component adaptation (Ou et al. 2004; Ren et al. 2005). Three additional learning statistics have also been set up, which include a predicted residual statistic (Xu and Yang 2000; Yang and Gao 2006a), a variance component ratio statistic from both the measurements and the predicted states (Yang and Xu 2003) and a velocity discrepancy between the predicted velocity from the kinematic model and the velocity evaluated from the measurements (Cui and Yang 2006).

Two optimal adaptive factors are established, which satisfy the conditions that the theoretical uncertainty of the predicted state outputted from the adaptive filtering equals or nearly equals its actual estimated uncertainty, and/or the theoretical uncertainty of the predicted residual vector equals or nearly equals its actual estimated uncertainty (Yang and Gao 2006b). Furthermore, an adaptively robust filter with classified adaptive factors is developed (Cui and Yang 2006) which is more effective in tracking the disturbances of the vehicle movements. And an adaptively robust filter with multi-adaptive factors is also set up (Yang et al. 2006), which is more general in theory and contains the adaptively robust filters with single adaptive factor and classified adaptive factors.

The new adaptively robust filter is more or less based on the assumption that the measurements at each epoch should be reliable. If it is not true, then the detection and identification and adaptation procedure can be employed (Teunissen 1990) or a robust Kalman filter can be applied (Koch and Yang 1998; Schaffrin 1991, pp. 32–34; Yang 1991, 1997; Zhou et al. 1997).

2.2 The Principle of Adaptively Robust Kalman Filtering

Let the linear dynamic system be given by

$$\mathbf{X}_k = \mathbf{\Phi}_{k,k-1}\mathbf{X}_{k-1} + \mathbf{W}_k, \tag{1}$$

where $\mathbf{X}_k$ denotes $m \times 1$ state vector at epoch t_k, $\mathbf{\Phi}_{k,k-1}$ the $u \times u$ transition matrix and $\mathbf{W}_k$ the state noise vector. The observational model at epoch t_k reads

$$\mathbf{L}_k = \mathbf{A}_k\mathbf{X}_k + \mathbf{e}_k, \tag{2}$$

where $\mathbf{L}_k$ represents $n_k \times 1$ observation vector, $\mathbf{A}_k$ the $n_k \times m$ design matrix and $\mathbf{e}_k$ the observational noise vector. Let the covariance matrices of $\mathbf{W}_k$ and $\mathbf{e}_k$ be taken as $\mathbf{\Sigma}_{\mathbf{W}_k}$ and $\mathbf{\Sigma}_k$, respectively, and $\mathbf{W}_k, \mathbf{W}_j, \mathbf{e}_k$ and $\mathbf{e}_j$ be mutually uncorrelated and meet:

$$E(\mathbf{e}_k) = 0, E(\mathbf{W}_k) = 0, \tag{3}$$

$$E(\mathbf{e}_k\mathbf{e}_k^T) = \mathbf{\Sigma}_k, E(\mathbf{W}_k\mathbf{W}_k^T) = \mathbf{\Sigma}_{\mathbf{w}_k}. \tag{4}$$

Assume that the residual vector is denoted by $\mathbf{V}_k$ and the predicted state vector is $\overline{\mathbf{X}}_k$; then the error equation and the predicted state vector are

$$\mathbf{V}_k = \mathbf{L}_k\hat{\mathbf{X}}_k - \mathbf{L}_k, \tag{5}$$

$$\overline{\mathbf{X}}_k = \mathbf{\Phi}_{k,k-1}\hat{\mathbf{X}}_{k-1}, \tag{6}$$

with

$$\mathbf{\Sigma}_{\overline{\mathbf{X}}_k} = \mathbf{\Phi}_{k,k-1}\mathbf{\Sigma}_{\hat{\mathbf{X}}_{k-1}}\mathbf{\Phi}^{\mathrm{T}}_{k,k-1} + \mathbf{\Sigma}_{\mathbf{w}_k}, \tag{7}$$

where $\hat{\mathbf{X}}_k$ and $\hat{\mathbf{X}}_{k-1}$ are the estimated state vectors at epochs $\mathrm{t_k}$ and $\mathrm{t_{k-1}}$.

By using the least squares principle,

$$\mathbf{V}_k^T\mathbf{P}_k\mathbf{V}_k + (\hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k)^T\mathbf{P}_{\overline{\mathbf{X}}_k}(\hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k) = \min, \tag{8}$$

where $\mathbf{P}_k = \mathbf{\Sigma}_k^{-1}$ and $\mathbf{P}_{\overline{\mathbf{X}}_k} = \mathbf{\Sigma}_{\overline{\mathbf{X}}_k}^{-\mathbf{1}}$ are the weight matrices of $\mathbf{L}_k$ and $\overline{\mathbf{X}}_k$, respectively; we obtain the estimator of the standard Kalman filter (Koch and Yang 1998):

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_\mathbf{k}})^{-1} (\mathbf{P}_{\overline{\mathbf{X}}_\mathbf{k}} \overline{\mathbf{X}}_k + \mathbf{A}_k^T \mathbf{P}_k L_k). \tag{9}$$

Or equivalently

$$\hat{\mathbf{X}}_k = \overline{\mathbf{X}}_k + \mathbf{K}_k (\mathbf{L}_k - \mathbf{A}_k \overline{\mathbf{X}}_k), \tag{10}$$

with

$$\mathbf{K}_k = \mathbf{\Sigma}_{\bar{\mathrm{X}}_k} \mathbf{A}_k^T (\mathbf{A}_k \mathbf{\Sigma}_{\overline{\mathrm{X}}_k} \mathbf{A}_k^T + \mathbf{\Sigma}_k)^{-\mathbf{1}}, \tag{11}$$

$$\mathbf{\Sigma}_{\hat{\mathbf{X}}_k} = [\mathbf{I} - \mathbf{K}_k \mathbf{A}_k] \mathbf{\Sigma}_{\overline{\mathbf{X}}_k}. \tag{12}$$

Changing a little bit of the score function of (8), like

$$\sum_{i=1}^{n_k} \mathrm{p}_i \rho(\mathrm{v}_i) + \alpha_k (\hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k)^T \mathbf{P}_{\overline{\mathbf{X}}_k} (\hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k) = \min, \tag{13}$$

where ρ is a convex and continuous function, α_k is an adaptive factor with values in $0 < \alpha_k \leq 1$, we get the estimator of the adaptive filter (Yang et al. 2001a):

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1} (\alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k} \overline{\mathbf{X}}_k + \mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{L}_k), \tag{14}$$

or equivalently (Yang et al. 2001b; Xu 2007)

$$\hat{\mathbf{X}}_k = \overline{\mathbf{X}}_k + \bar{\mathbf{K}}_k (\mathbf{L}_k - \mathbf{A}_k \overline{\mathbf{X}}_k), \tag{15}$$

where $\bar{\mathbf{K}}_k$ is an adaptive gain matrix:

$$\bar{\mathbf{K}}_k = \frac{1}{\alpha_k} \mathbf{\Sigma}_{\bar{\mathrm{X}}_k} \mathbf{A}_k^T (\frac{1}{\alpha_k} \mathbf{A}_k \mathbf{\Sigma}_{\bar{\mathrm{X}}_k} \mathbf{A}_k^T + \bar{\mathbf{\Sigma}}_k)^{-\mathbf{1}}. \tag{16}$$

The posterior covariance matrix of the estimated state vector is

$$\mathbf{\Sigma}_{\hat{\mathrm{X}}_k} = (\mathbf{I} - \bar{\mathbf{K}}_k \mathbf{A}_k) \mathbf{\Sigma}_{\bar{\mathrm{X}}_k} / \alpha_k. \tag{17}$$

With the variations of adaptive factor α_k and the equivalent weight matrix, the adaptively robust filter will change into various estimators.

Case 1: if $\alpha_k = 0$ and $\bar{\mathbf{\Sigma}}_k = \mathbf{\Sigma}_k$ or $\bar{\mathbf{P}}_k = \mathbf{P}_k$, then

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k)^{-1} \mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k, \tag{18}$$

which is an LS estimator by using only the new measurements at epoch t_k. This estimator is suitable to the case that the measurements are not contaminated by outliers and the updated parameters are biased so much that the information of updated parameters should be forgotten completely.

Case 2: if $\alpha_k = 1$ and $\bar{\mathbf{\Sigma}}_k = \mathbf{\Sigma}_k$, then the standard Kalman filter estimators (8), (9) and (10) are obtained.

Case 3: if α_k changes between 0 and 1, and $\bar{\mathbf{P}}_k = \mathbf{P}_k$, then

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\bar{\mathbf{X}}_k})^{-1} (\mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k + \alpha_k \mathbf{P}_{\bar{\mathbf{X}}_k} \overline{\mathbf{X}}_k), \tag{19}$$

which is an adaptive LS estimator of Kalman filter. It balances the contribution of the updated parameters and the measurements. The only difference between (17) and (8) is the weight matrix of $\mathbf{L}_k$. The former uses the equivalent weight matrix and the latter uses the original weight matrix of $\mathbf{L}_k$.

Case 4: if $\alpha_k = 0$, then we obtain

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{A}_k)^{-1} \mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{L}_k, \tag{20}$$

which is a robust estimator by using only the new measurements at epoch t_k.

Case 5: if $\alpha_k = 1$, then

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1} (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{L}_k + \mathbf{P}_{\overline{\mathbf{X}}_k} \overline{\mathbf{X}}_k), \tag{21}$$

which is an M-LS filter estimator (Yang 1997).

The relations of the adaptively robust filter with other estimators are shown in Fig. 2.1, in which ARF denotes the adaptively robust filter.

If the covariance matrices of the measurement vector $\mathbf{L}_k$ and the predicted state vector $\overline{\mathbf{X}}_k$ are evaluated by Sage windowing method (see Deng 2003), denoted as $\hat{\mathbf{\Sigma}}_k$ and $\hat{\mathbf{\Sigma}}_{\overline{\mathbf{X}}_k}$, respectively, that is,

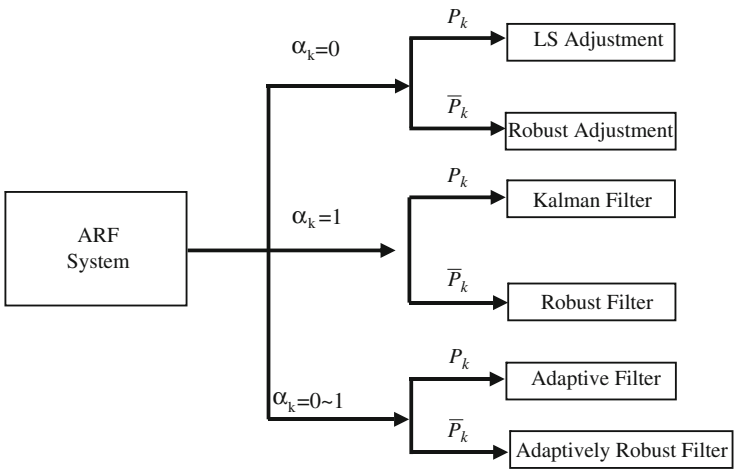


Fig. 2.1 Adaptively robust filter

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{m}\sum_{j=0}^{m} \bar{\mathbf{V}}_{k-j}\bar{\mathbf{V}}_{k-j}^T - \mathbf{A}_k \boldsymbol{\Sigma}_{\overline{\mathbf{X}}_k} \mathbf{A}_k^T, \tag{22}$$

$$\hat{\boldsymbol{\Sigma}}_{\Delta X} = \frac{1}{m}\sum_{j=0}^{m} \Delta \mathbf{X}_{k-j} \Delta \mathbf{X}_{k-j}^T, \tag{23}$$

where

$$\bar{\mathbf{V}}_k = \mathbf{A}_k \overline{\mathbf{X}}_k - \mathbf{L}_k, \tag{24}$$

$$\Delta \mathbf{X}_k = \hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k, \tag{25}$$

then the adaptively robust filter can include the adaptive Sage filter.

2.3 Properties of the Adaptive Kalman Filter

2.3.1 Difference of State Estimate

Rewriting (9) and (14) as normal equations, and only considering the least squares situation, that is, $\bar{\mathbf{P}}_k = \mathbf{P}_k$, we have

$$(\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})\hat{\mathbf{X}}_k = \mathbf{P}_{\overline{\mathbf{X}}_k}\overline{\mathbf{X}}_k + \mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k, \tag{26}$$

$$(\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \boldsymbol{\alpha}_k \mathbf{P}_{\overline{\mathbf{X}}_k})\hat{\mathbf{X}}_{\mathbf{ad}} = \boldsymbol{\alpha}_k \mathbf{P}_{\overline{\mathbf{X}}_k}\overline{\mathbf{X}}_k + \mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k, \tag{27}$$

where $\hat{\mathbf{X}}_k$ and $\hat{\mathbf{X}}_{\text{ad}}$ denote the state estimates by using standard Kalman filter and adaptive filter, respectively. Let

$$\hat{\mathbf{X}}_k - \hat{\mathbf{X}}_{\mathbf{ad}} = \boldsymbol{\delta}\hat{X}_k, \tag{28}$$

then (27) is changed into

$$(\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k})\hat{\mathbf{X}}_k - (\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k})\boldsymbol{\delta}\hat{\mathbf{X}}_k = \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k}\overline{\mathbf{X}}_k + \mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k. \tag{29}$$

Subtracting (29) from (26), we have

$$(1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}\hat{\mathbf{X}}_k + (\mathbf{A}_k^T \mathbf{P}_k \mathbf{L}_k + \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k})\boldsymbol{\delta}\hat{\mathbf{X}}_k = (1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}\overline{\mathbf{X}}_k \tag{30}$$

or

$$(\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\overline{\mathbf{X}}_k})\boldsymbol{\delta}\hat{\mathbf{X}}_k = (1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}(\bar{\mathbf{X}}_k - \hat{\mathbf{X}}_k). \tag{31}$$

Denote

$$\overline{\mathbf{X}}_k - \hat{\mathbf{X}}_{\mathbf{ad}} = \boldsymbol{\delta}\bar{X}_k. \tag{32}$$

Factually, $\boldsymbol{\delta}\bar{\mathbf{X}}_k$ can be looked as the bias of the predicted state, and (31) is changed into

$$(\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})\boldsymbol{\delta}\hat{\mathbf{X}}_k = (1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}(\boldsymbol{\delta}\bar{\mathbf{X}}_k + \hat{\mathbf{X}}_{\mathbf{ad}} - \hat{\mathbf{X}}_k). \tag{33}$$

Finally we have

$$(\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})\boldsymbol{\delta}\hat{\mathbf{X}}_k = (1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}\boldsymbol{\delta}\bar{\mathbf{X}}_k, \tag{34}$$

$$\delta\hat{\mathbf{X}}_k = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(1-\alpha_k)\mathbf{P}_{\overline{\mathbf{X}}_k}\boldsymbol{\delta}\bar{\mathbf{X}}_k. \tag{35}$$

It is easy to see that if $\alpha_k = 1$, then $\boldsymbol{\delta}\hat{\mathbf{X}}_k = 0$, that is, $\hat{\mathbf{X}}_{\mathbf{ad}} = \hat{\mathbf{X}}_k$; in this case the state estimate of the adaptive filter is equivalent to that of the standard Kalman filter; if $\alpha_k = 0$, then

$$\hat{\mathbf{X}}_{\mathbf{ad}} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k)^{-\mathbf{1}}\mathbf{A}_k^T\mathbf{P}_k\mathbf{L}_k. \tag{36}$$

In this case, $\hat{\mathbf{X}}_{\mathbf{ad}}$ is equivalent to the estimate of the least squares estimation not considering the information of the state equation, and the error of the predicted state $\overline{\mathbf{X}}_k$ will not affect the updated estimate of the state.

2.3.2 The Expectation of the State Estimate of the Adaptive Filter

Considering that the observational vector is unbiased, that is $\mathrm{E}\mathbf{L}_k = \mathbf{A}_k\mathbf{X}_k$, then the expectation of $\hat{\mathbf{X}}_{\mathbf{ad}}$ from (27) is

$$\mathrm{E}\hat{\mathbf{X}}_{\mathrm{ad}} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(\alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k}\mathrm{E}\overline{\mathbf{X}}_k + \mathbf{A}_k^T\mathbf{P}_k\mathbf{L}_k\mathbf{X}_k). \tag{37}$$

If the predicted state vector $\overline{\mathbf{X}}_k$ is also unbiased, that is $\mathrm{E}\overline{\mathbf{X}}_k = \mathbf{X}_k$, then

$$\mathrm{E}\hat{\mathbf{X}}_{\mathbf{ad}} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})\mathbf{X}_k = \mathbf{X}_k. \tag{38}$$

It is obvious that if the observational vector $\mathbf{L}_k$ and the predicted state vector $\overline{\mathbf{X}}_k$ are not affected by abnormal biases, then the estimate of the adaptive filter is unbiased.

If the predicted state vector $\overline{\mathbf{X}}_k$ is biased, and the biased vector is denoted by $\mathbf{b}_{\overline{\mathbf{X}}_k}$, that is,

$$\mathrm{E}\overline{\mathbf{X}}_k = \mathbf{X}_k + \mathbf{b}_{\overline{\mathbf{X}}_k} \neq \mathbf{X}_k, \tag{39}$$

then

$$\mathrm{E}\hat{\mathbf{X}}_{\mathbf{ad}} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(\alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{X}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k} + \mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k\mathbf{X}_k), \tag{40}$$

then we have

$$\mathrm{E}\hat{\mathbf{X}}_{\mathbf{ad}} = \mathbf{X}_k + (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k} \tag{41}$$

or

$$\mathbf{b}_{\hat{\mathbf{X}}_{\mathbf{ad}}} = \mathrm{E}\hat{\mathbf{X}}_{\mathbf{ad}} - \mathbf{X}_k = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k}. \tag{42}$$

Equation (42) gives the influence of the bias $\mathbf{b}_{\overline{\mathbf{X}}_k}$ of $\overline{\mathbf{X}}_k$ on the bias $\mathbf{b}_{\hat{\mathbf{X}}_{\mathbf{ad}}}$ of the estimate expectation of the adaptive filter. Equation (42) also tells us that $\mathbf{b}_{\hat{\mathbf{X}}_{\mathrm{ad}}}$ changes with the variations of α_k and $\mathbf{b}_{\overline{\mathbf{X}}_k}$, especially when $\mathbf{b}_{\overline{\mathbf{X}}_k}$ is beyond a particular region; α_k tends to zero and, in this case, $\mathbf{b}_{\hat{\mathbf{X}}_{\mathbf{ad}}}$ also tends to zero. In other words, the bias of the state estimate of the adaptive filter is controlled by the adaptive factor α_k. Usually the larger the bias $\mathbf{b}_{\overline{\mathbf{X}}_k}$ of the predicted state, the smaller the adaptive factor α_k.

Similarly, when $\mathrm{E}\overline{\mathbf{X}}_k \neq \mathbf{X}_k$, the expectation of the state estimate of the standard Kalman filter is

$$\mathrm{E}\hat{\mathbf{X}}_k = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{X}_k + \mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k} + \mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k\mathbf{X}_k) \tag{43}$$

or

$$\mathrm{E}\hat{\mathbf{X}}_k = \mathbf{X}_k + (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k}. \tag{44}$$

Let

$$\mathbf{b}_{\hat{\mathbf{X}}_k} = \mathrm{E}\hat{\mathbf{X}}_k - \mathbf{X}_k, \left(\text{bias of } \hat{\mathbf{X}}_k\right). \tag{45}$$

Then

$$\mathbf{b}_{\hat{\mathbf{X}}_k} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\mathbf{P}_{\overline{\mathbf{X}}_k}\mathbf{b}_{\overline{\mathbf{X}}_k}. \tag{46}$$

2.3.3 Posterior Precision Evaluation

By applying the variance propagation law we obtain the covariance matrices of the estimated state vectors of the adaptive filter and the standard Kalman filter, respectively, as

$$\Sigma_{\hat{\mathbf{X}}_{\mathrm{ad}}} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}(\alpha_k^2\mathbf{P}_{\overline{\mathbf{X}}_k} + \mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k)(\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \alpha_k\mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\hat{\sigma}_{\mathrm{ad}}^2, \tag{47}$$

$$\Sigma_{\hat{\mathbf{X}}_k} = (\mathbf{A}_k^T\mathbf{P}_k\mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k})^{-1}\hat{\sigma}^2, \tag{48}$$

in which

$$\hat{\sigma}_{\text{ad}}^2 = \frac{\alpha_k \delta\hat{\mathbf{X}}_{\mathbf{ad}}^T \mathbf{P}_{\overline{\mathbf{X}}_k} \delta\hat{\mathbf{X}}_{\mathbf{ad}} + \mathbf{V}_{\mathbf{ad}}^T \mathbf{P}_k \mathbf{V}_{\mathbf{ad}}}{r_k}, \tag{49}$$

$$\hat{\sigma}^2 = \frac{\delta\hat{\mathbf{X}}_k^T \mathbf{P}_{\overline{\mathbf{X}}_k} \delta\hat{\mathbf{X}}_k + \mathbf{V}_k^T \mathbf{P}_k \mathbf{V}_k}{r_k}, \tag{50}$$

where r_k is redundant number of the observations, $\mathbf{V}_{\mathbf{ad}}$ and $\mathbf{V}_k$ are residual vectors of the observations with respect to $\hat{\mathbf{X}}_{\mathbf{ad}}$ and $\hat{\mathbf{X}}_k$, respectively.

In practice, the posterior precision is evaluated by the mean-square error (Xu and Rummel 1994), that is,

$$\text{MSE}(\hat{\mathbf{X}}_{\mathbf{ad}}) = \text{E}(\hat{\mathbf{X}}_{\mathbf{ad}} - \mathbf{X}_k)^T(\hat{\mathbf{X}}_{\mathbf{ad}} - \mathbf{X}_k), \tag{51}$$

where $\mathbf{X}_k$ denotes the true value of the state vector. Changing (51) as

$$\text{MSE}(\hat{\mathbf{X}}_{\mathbf{ad}}) = \text{E}(\hat{\mathbf{X}}_{\mathbf{ad}} - \text{E}\hat{\mathbf{X}}_{\mathbf{ad}} + \text{E}\hat{\mathbf{X}}_{\mathbf{ad}} - \mathbf{X}_k)^T(\hat{\mathbf{X}}_{\mathbf{ad}} - \text{E}\hat{\mathbf{X}}_{\mathbf{ad}} + \text{E}\hat{\mathbf{X}}_{\mathbf{ad}} - \mathbf{X}_k), \tag{52}$$

letting

$$\hat{\mathbf{X}}_{\text{ad}} - \text{E}\hat{\mathbf{X}}_{\text{ad}} = \mathbf{e}_{\hat{\mathbf{X}}_{\text{ad}}} \left(\text{true error vector of } \hat{\mathbf{X}}_{\text{ad}}\right) \tag{53}$$

and considering

$$\text{E}\mathbf{X}_k = \mathbf{X}_k, \text{E}\left(\hat{\mathbf{X}}_{\text{ad}} - \text{E}\hat{\mathbf{X}}_{\text{ad}}\right) = 0, \tag{54}$$

we get

$$\text{MSE}(\hat{\mathbf{X}}_{\text{ad}}) = \text{trE}(\mathbf{e}_{\hat{\mathbf{X}}_{\mathbf{ad}}} \cdot \mathbf{e}_{\hat{\mathbf{X}}_{\text{ad}}}^T) + \left|\left|\mathbf{b}_{\hat{\mathbf{X}}_{\text{ad}}}\right|\right|^2. \tag{55}$$

On the other hand, we have

$$\text{E}\left(\mathbf{e}_{\hat{\mathbf{x}}_{\text{ad}}} \cdot \mathbf{e}_{\hat{\mathbf{x}}_{\text{ad}}}^T\right) = \sum\nolimits_{\hat{\mathbf{x}}_{\text{ad}}} \left(\text{covariance matrix of } \hat{\mathbf{x}}_{\text{ad}}\right), \tag{56}$$

which then yields

$$\text{MSE}(\hat{\mathbf{X}}_{\mathbf{ad}}) = \text{tr}\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{\mathbf{ad}}} + \alpha_k \left\|(\mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \alpha_k P_{\bar{X}_k})^{-1} \mathbf{P}_{\bar{X}_k} (\text{E}\overline{\mathbf{X}}_k - \mathbf{X}_k)\right\|^2. \tag{57}$$

It is seen from (57) that

1. If the predicted state vector $\overline{\mathbf{X}}_k$ is unbiased, that is $\mathbf{b}_{\mathbf{X}_k} = 0$, then $\hat{\mathbf{X}}_{\mathbf{ad}}$ is unbiased, or $\mathbf{b}_{\hat{\mathbf{X}}_{\mathbf{ad}}} = 0$; in this case, the MSE of $\hat{\mathbf{X}}_{\mathbf{ad}}$ is the trace of its covariance matrix,

$$\text{MSE}(\hat{\mathbf{X}}_{\mathbf{ad}}) = \text{tr}(\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{\mathbf{ad}}}). \tag{58}$$

2. If $\overline{\mathbf{X}}_k$ is biased, but $\alpha_k = 0$, (57) is still valid, that is the adaptive factor α_k controls the bias of the outputs of the adaptive filter.

Therefore, when the predicted state $\overline{\mathbf{X}}_k$ has any abnormal bias due to some sudden disturbance of the vehicle, the adaptive factor α_k will be decreased, which leads to the bias $\mathbf{b}_{\hat{X}_{\text{ad}}}$ of the state estimate of the adaptive filter to decrease, and the mean-square error of $\hat{\mathbf{X}}_{\mathbf{ad}}$ tends to the trace of the covariance matrix of $\hat{\mathbf{X}}_{\mathbf{ad}}$.

In conclusion, the differences between the adaptive filter and the standard Kalman filter depend on the adaptive factor α_k. When the predicted states are accurate, then α_k tends to 1, and the state differences estimated from the adaptive filter and the standard Kalman filter are small.

The unbiasedness of the estimated state vector outputted by adaptively filter is controlled by the adaptive factor α_k; if α_k tends to zero, $\hat{\mathbf{X}}_{\mathbf{ad}}$ is unbiased.

The MSE of the estimated state vector outputted by the adaptive filter is also controlled by the adaptive factor α_k; if α_k tends to zero, the MSE of $\hat{\mathbf{X}}_{\mathbf{ad}}$ tends to the trace of its covariance matrix.

The robustness of the adaptive filter outputs has been described in Yang et al. (2001a, b) and Yang and Xu (2003). It has been demonstrated by theory and practical experiments that the adaptive factor plays significant roles in controlling the influences of the outlying disturbances of the dynamical information on the estimated state vector and its MSE.

2.4 Three Kinds of Learning Statistics

2.4.1 Learning Statistic Constructed by State Discrepancy

In the beginning of the development of the adaptively robust Kalman filter, a learning statistic of the kinematic model errors was constructed by using the difference between the state estimated from measurements and that predicted from the kinematic model at epoch t_k (Yang et al. 2001a, b). If the number of measurements at epoch t_k is larger than that of the state components, then the estimated state vector is obtained by using measurement vector L_k, based on the robust estimation principle, that is,

$$\tilde{\mathbf{X}}_k = (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{A}_k)^{-1} \mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{L}_k, \tag{59}$$

where $\bar{\mathbf{P}}_k$ denotes the equivalent weight matrix of $\mathbf{L}_k$, which can be obtained by the Huber function (Huber 1981), three-segment functions (Yang 1994, 1999; Yang et al. 2002a, b; Zhou 1989),etc.

The discrepancy between $\tilde{\mathbf{X}}_k$ from (59) and $\overline{\mathbf{X}}_k$ from (6) can be measured by

$$\left\|\tilde{\mathbf{X}}_k - \overline{\mathbf{X}}_k\right\| = \left(\Delta\tilde{\mathbf{X}}_{k_1}^2 + \Delta\tilde{\mathbf{X}}_{k_2}^2 + \cdots + \Delta\tilde{\mathbf{X}}_{k_m}^2\right)^{\frac{1}{2}}. \tag{60}$$

Then the learning statistic expressed by the state discrepancy is set up:

$$\left|\Delta\tilde{\mathbf{X}}_k\right| = \frac{\left\|\tilde{\mathbf{X}}_k - \overline{\mathbf{X}}_k\right\|}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{\overline{\mathbf{X}}_k})}}, \tag{61}$$

where "tr" denotes trace.

It is noted that (1) the number of measurements at computation epoch should be larger than the number of state components, otherwise the statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ cannot be constructed; (2) the estimated state vector $\tilde{\mathbf{X}}_k$ should be accurate, otherwise the statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ cannot reflect the kinematic model errors; and (3) the learning statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ can only reflect the integrated error of the kinematic model; any disturbing of the components of the predicted state vector is treated as the whole state outlier.

2.4.2 *Learning Statistic Constructed by Predicted Residual Vector*

If the measurement vector $\mathbf{L}_k$ is reliable, then the predicted residual vector $\bar{\mathbf{V}}_k$ will reflect the error of predicted state vector $\overline{\mathbf{X}}_k$. A learning statistic constructed by the predicted residual vector is (Xu and Yang 2000; Yang and Gao 2006a)

$$\Delta\bar{\mathbf{V}}_k = \left(\frac{\bar{\mathbf{V}}_k^T\bar{\mathbf{V}}_k}{\mathrm{tr}(\boldsymbol{\Sigma}_{\bar{\mathbf{V}}_k})}\right)^{\frac{1}{2}}. \tag{62}$$

If there are n_k measurements at epoch t_k, then $\Delta\bar{\mathbf{V}}_k$ can be expressed as

$$\Delta\bar{\mathbf{V}}_k = \left(\frac{\sum_{i=1}^{n_k}\bar{\mathbf{V}}_{k_i}^2}{\sum_{i=1}^{n_k}\sigma_{\bar{\mathbf{V}}_{k_i}}^2}\right)^{\frac{1}{2}}. \tag{63}$$

It is noted that (1) using the learning statistic constructed by predicted residual $\Delta\bar{\mathbf{V}}_k$, we do not need to evaluate the state vector before filtering; (2) it is not necessary that the number of measurements be larger than that of state components; and (3) $\Delta\bar{\mathbf{V}}_k$ contains more measurement error influence than the state discrepancy statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$.

2.4.3 Learning Statistic Constructed by the Ratio of Variance Components

If $\mathbf{L}_k$ and $\overline{\mathbf{X}}_k$ are treated as two groups of observations at epoch t_k, their variance components should reflect their accuracies. Thus we can construct a new learning statistic by the ratio of the variance components to judge the kinematic model reliability. The Helmert variance component for $\mathbf{L}_k$ and $\overline{\mathbf{X}}_k$ is respectively expressed as (Koch 2000; Koch and Kusche 2002)

$$\hat{\sigma}_{0k}^2 = \frac{\mathbf{V}_k^T \mathbf{P}_k \mathbf{V}_k}{r_k - \mathrm{tr}(\mathbf{N}^{-1}\mathbf{N}_k)}, \tag{64}$$

$$\hat{\sigma}_{0\bar{X}_k}^2 = \frac{\mathbf{V}_{\overline{\mathbf{X}}_k}^T \mathbf{P}_{\overline{\mathbf{X}}_k} \mathbf{V}_{\overline{\mathbf{X}}_k}}{m_k - \mathrm{tr}(\mathbf{N}^{-1}\mathbf{P}_{\overline{\mathbf{X}}_k})}, \tag{65}$$

where $\hat{\sigma}_{0\,k}^2$ and $\hat{\sigma}_{0\overline{\mathbf{X}}_k}^2$ denote the estimates of variance components of $\mathbf{L}_k$ and $\overline{\mathbf{X}}_k$, respectively, n_k is the number of measurements at epoch t_k, m_k is the number of predicted parameters of the state vector, $\mathbf{V}_k$ is the residual vector of $\mathbf{L}_k$ expressed by (5) and $\mathbf{V}_{\overline{\mathbf{X}}_k}$ is the residual vector of $\overline{\mathbf{X}}_k$, that is,

$$\mathbf{V}_{\overline{\mathbf{X}}_k} = \hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k = \hat{\mathbf{X}}_k - \mathbf{\Phi}_{k,k-1}\hat{\mathbf{X}}_{k-1} \tag{66}$$

and

$$\mathbf{N}_k = \mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k, \mathbf{N} = \mathbf{N}_k + \mathbf{P}_{\overline{\mathbf{X}}_k} = \mathbf{A}_k^T \mathbf{P}_k \mathbf{A}_k + \mathbf{P}_{\overline{\mathbf{X}}_k}. \tag{67}$$

The approximate estimates of the Helmert variance components $\hat{\sigma}_{0\,k}^2$ and $\hat{\sigma}_{0\overline{\mathbf{X}}_k}^2$ are

$$\sigma_{0\,k}^2 \approx \frac{\mathbf{V}_k^T \mathbf{P}_k \mathbf{V}_k}{n_k} \tag{68}$$

and

$$\sigma_{0\bar{X}_k}^2 \approx \frac{\mathbf{V}_{\overline{\mathbf{X}}_k}^T \mathbf{P}_{\overline{\mathbf{X}}_k} \mathbf{V}_{\overline{\mathbf{X}}_k}}{m_k}. \tag{69}$$

The ratio of $\hat{\sigma}_{0\overline{\mathbf{X}}_k}^2$ and $\hat{\sigma}_{0\,k}^2$ is defined as the learning statistic

$$S_k = \frac{\sigma_{0\overline{\mathbf{X}}_k}^2}{\sigma_{0\,k}^2} \approx \frac{\mathbf{V}_{\overline{\mathbf{X}}_k}^T \mathbf{P}_{\overline{\mathbf{X}}_k} \mathbf{V}_{\overline{\mathbf{X}}_k}}{m_k \sigma_{0\,k}^2}. \tag{70}$$

It is noted that (1) the computation of the learning statistic S_k should have redundant observations, otherwise the learning statistic will not reliably reflect the model

errors; (2) $\mathbf{V}_k$ and $\mathbf{V}_{\overline{\mathbf{X}}_k}$ should correspond to the same estimated state vector $\hat{\mathbf{X}}_k$; and (3) the computation burden is heavier than the other two learning statistics, $\left|\Delta\tilde{\mathbf{X}}_k\right|$ and $\Delta\bar{\mathbf{V}}_k$, if the iterative computation procedures are applied.

2.4.4 Learning Statistic Constructed by Velocity

Based on the robust estimate $\tilde{\mathbf{X}}_k$ of the position parameter vector from the measurements, the estimate $\hat{\mathbf{X}}_{k-1}$ of the state estimate at epoch t_{k-1} and the sample interval $t_k - t_{k-1}$, we obtain the predicted velocity vector (Cui and Yang 2006)

$$\tilde{\dot{\mathbf{X}}}_k = \frac{\tilde{\mathbf{X}}_k - \hat{\mathbf{X}}_{k-1}}{t_k - t_{k-1}}. \tag{71}$$

Then the learning statistic for judging the kinematic model disturbing corresponding to the predicted velocity information is constructed as

$$\Delta\tilde{\dot{\mathbf{X}}}_k = \left\|\tilde{\dot{\mathbf{X}}}_k - \bar{\dot{\mathbf{X}}}_k\right\| \Big/ \sqrt{\mathrm{tr}(\Sigma_{\bar{\dot{\mathbf{X}}}_k})}, \tag{72}$$

where $\bar{\dot{\mathbf{X}}}_k$ denotes the predicted velocity vector from the kinematic model and $\Sigma_{\bar{\dot{\mathbf{X}}}_k}$ is its covariance matrix.

It is noted that (1) if $\Delta\tilde{\dot{\mathbf{X}}}_k$ is significant outlying, then it indicates that the predicted velocity is outlying or that the kinematic model has large errors and (2) the computation of the learning statistic $\Delta\tilde{\dot{\mathbf{X}}}_k$ should also have redundant observations, otherwise $\tilde{\mathbf{X}}_k$ cannot be obtained.

2.5 Four Kinds of Adaptive Factors

2.5.1 Adaptive Factor by Three-Segment Function

An adaptive factor α_k of a three-segment function is combined by three parts, that is, if a learning statistic is smaller than a particular criterion, then the adaptive factor α_k is equal to 1, if the learning statistic is significantly outlying, then the adaptive factor α_k is equal to 0, otherwise α_k decreases with the statistic growing. We employ the learning statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ as an example to express the adaptive factor (Yang et al. 2001a):

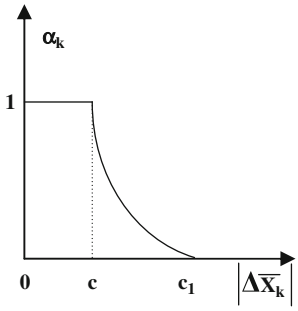


Fig. 2.2 Adaptive factor of three-segment function

$$\alpha_k = \begin{cases} 1, & \left|\Delta\tilde{X}_k\right| \le c_0, \\ \frac{c_0}{\left|\Delta\tilde{X}_k\right|}\left(\frac{c_1 - \left|\Delta\tilde{X}_k\right|}{c_1 - c_0}\right)^2, & c_0 < \left|\Delta\tilde{X}_k\right| \le c_1, \\ 0, & \left|\Delta\tilde{X}_k\right| > c_1, \end{cases} \tag{73}$$

where c_0 and c_1 are two criterion constants, with practical values of $c_0 = 1.0 - 1.5$ and $c_1 = 3.0 - 4.5$.

Obviously, if the value of $\left|\Delta\tilde{\mathbf{X}}_k\right|$ increases, α_k decreases. α_k changes between [0, 1] (see Fig. 2.2). This kind of adaptive factor is a redescending function, because α_k descends to zero when the statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ is larger than the rejection boundary c_1.

2.5.2 Adaptive Factor by Two-Segment Function

We still employ the learning statistic $\left|\Delta\tilde{\mathbf{X}}_k\right|$ as an example to express the two-segment adaptive factor (Yang et al. 2001b):

$$\alpha_k = \begin{cases} 1, & \left|\Delta\tilde{X}_k\right| \le c, \\ \frac{c}{\left|\Delta\tilde{X}_k\right|}, & \left|\Delta\tilde{X}_k\right| > c, \end{cases} \tag{74}$$

where c is a constant, the optimal value being 1.0 (Yang and Gao 2006a). It is a descending function; see Fig. 2.3.

2.5.3 Adaptive Factor by Exponential Function

An adaptive factor of exponential function is (Yang and Gao 2005)

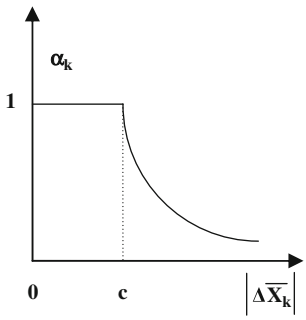


Fig. 2.3 Adaptive factor of two-segment function

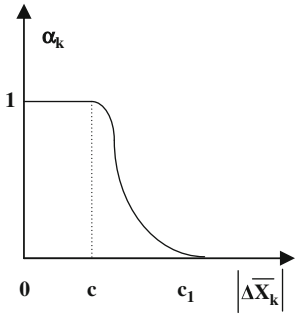


Fig. 2.4 Adaptive factor of exponential function

$$\alpha_k = \begin{cases} \mathbf{1}, & \left|\Delta\tilde{\mathbf{X}}_k\right| \le \mathbf{c}, \\ \mathbf{e}^{-(|\Delta\tilde{X}_k|-c)^2}, & \left|\Delta\tilde{\mathbf{X}}_k\right| > \mathbf{c}, \end{cases} \tag{75}$$

where c is the same as in (74). It is also a descending function; see Fig. 2.4.

2.5.4 Adaptive Factor by Zero and One

If a state parameter is normal then the adaptive factor equals 1, otherwise it equals 0 (Ou et al. 2004; Ren et al. 2005):

$$\alpha_k = \begin{cases} 1, & \left|\Delta\tilde{\mathbf{X}}_{\mathrm{ki}}\right| \le \mathrm{c}, \\ 0, & \left|\Delta\tilde{\mathbf{X}}_{\mathrm{ki}}\right| > \mathrm{c}, \end{cases} \tag{76}$$

where $\Delta\tilde{\mathbf{X}}_{\mathrm{ki}}$ is the ith component of the discrepancy state vector.

We can use another two learning statistics, $\Delta\bar{\mathbf{V}}_k$ from the predicted residual vector and S_k from the ratio of variance components of the predicted state and measurements, to construct the same kinds of adaptive factors.

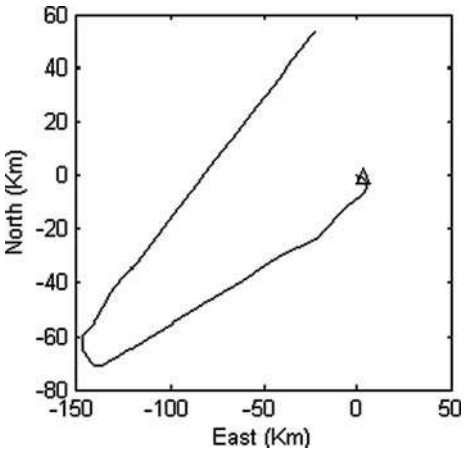


Fig. 2.5 Position relative to fixed receiver

2.5.5 Actual Computation and Analysis

The computation example is the same as that in Yang et al. (2001a). A data set was collected by using a Trimble 4000SSE receiver mounted on an aircraft with the reference receiver fixed at a site about 1 km from the initial aircraft location. After about 10 min of static tracking, the aircraft took off for a flight time of about 90 min; see Fig. 2.5.

In order to analyse the roles of the adaptive factors in adaptive filtering, the highly precise results from double-differenced carrier measurements were used as "true values" for comparing with the results from the code measurements. The constant-velocity model of the Kalman filtering was employed. The initial variances for positions, velocities and P2-code measurements were selected separately as $0.2\text{m}^2, 9 \times 10^{-5}\text{m}^2/\text{s}^2$and $1.0\ \text{m}^2$. The spectral density for velocities was chosen to be $0.01\ \text{m}^2/\text{s}^2$. The selected dynamic model covariance matrix was the same as that used in Schwarz et al. (1989), Yang et al. (2001a), Yang and Xu (2003) and Yang and Wen (2003).

The following four schemes were carried out:

- Scheme 1: Classical Kalman filtering (KF)
- Scheme 2: Adaptive Kalman filtering based on the three-segment function of the state discrepancy (AKF1)
- Scheme 3: Adaptive Kalman filtering based on the two-segment function of the state discrepancy (AKF2)
- Scheme 4: Adaptive Kalman filtering based on the exponential function of the state discrepancy (AKF3)

The computation results are shown in Figs. 2.6, 2.7, 2.8 and 2.9 and Table 2.1.

From the calculation results above, we find that

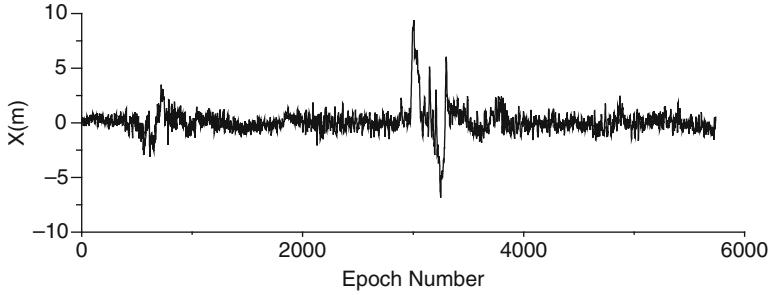
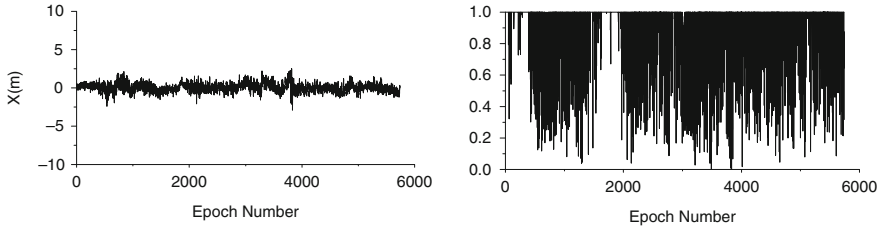


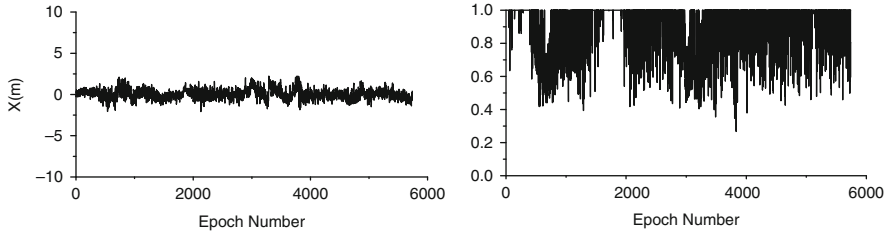
Fig. 2.6 Errors of classical Kalman filtering



a Errors of AKF1.

b Actual α_k determined by three segment function, where $c_0 = 1.5$, $c_1 = 4.5$

Fig. 2.7 (**a**) Errors of AKF1. (**b**) Actual α_k determined by three-segment function, where $c_0 = 1.5$ and $c_1 = 4.5$



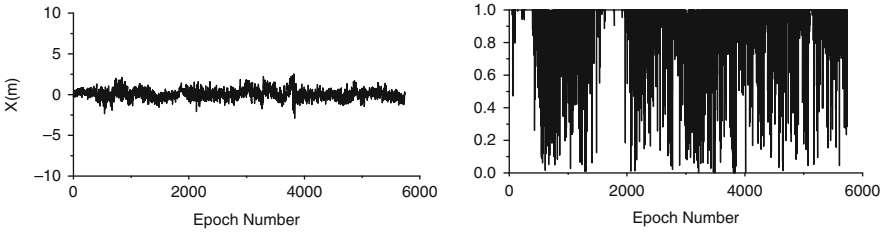
a Errors of AKF2.

b Actual α_k determined by two segment function, where $c = 1.5$

Fig. 2.8 (**a**) Errors of AKF2. (**b**) Actual α_k determined by two-segment function, where $c = 1.5$

Table 2.1 Comparison of RMS (unit: m)

	KF	AKF1	AKF2	AKF3
X	1.1630	0.5648	0.5948	0.5839
Y	1.5070	0.4438	0.5119	0.4766
Z	1.5455	0.7804	0.8201	0.8028



A Errors of AKF3.

B Actual α_k determined by exponential functiovn, where c = 1.5

Fig. 2.9 (**a**) Errors of AKF3. (**b**) Actual α_k determined by exponential function, where $c = 1.5$

(1) The influences of the disturbances of the fly around the epoch 1,000 and during the epoch 3,000 through 4,000 on the Kalman filtering are very significant; see Fig. 2.6. The RMS of the position are 1.1630, 1.5070 and 1.5455 m, respectively.
(2) The adaptive filtering based on the three kinds of adaptive factors gives reasonable results, and the influences of the disturbances of the fly are controlled.
(3) It is shown, by the theoretical and practical charting curves of the adaptive factors, that the result corresponding to the three-segment function is superior to those corresponding to the two-segment function and the exponential function. The reason is that the three-segment function decreases the adaptive factor quickly when the errors of the predicted state increase and gives the significant outlying predicted state zero factor; this kind of outlying predicted state from the kinematic model does not have any effect on the filtering results; see the second, third and fourth columns of Table 2.1 (AKF1, AKF2 and AKF3), respectively.

2.6 Comparison of Two Fading Filters and Adaptively Robust Filter

In order to control the influences of prior state errors or kinematic model errors on the present estimated state parameters, the fading filters, using the fading factors to restrict the memory length of Kalman filter and to make the most use of present measurements, were developed in the field of statistics as early as the 1960s and 1970s (Fagin 1964; Sorenson and Sacks 1971). The basic properties of the fading filter have been analysed, and the abilities in controlling the influences of the kinematic model errors on the state parameter estimates and the possibly existing problems in the practical applications have been discussed, respectively (Yang and Gao 2006c).

2.6.1 Principles of Two Kinds of Fading Filters

Assume that the estimated state vector and the residual vector are, respectively, $\hat{\mathbf{X}}_{k-1}$ and $\mathbf{V}_{\hat{\mathbf{X}}_{k-1}}$ at epoch t_{k-1}; the corresponding re-estimated state vector of $\mathbf{X}_{k-1}$ by using new measurement vector $\mathbf{L}_k$ at epoch t_k is denoted by $\hat{\hat{\mathbf{X}}}_{k-1}$; then the corresponding error equations are, respectively,

$$\mathbf{V}_{\hat{\mathbf{X}}_{k-1}} = \hat{\hat{\mathbf{X}}}_{k-1} - \hat{\mathbf{X}}_{k-1}, \tag{77}$$

$$\hat{\mathbf{w}}_{k-1} = \hat{\mathbf{X}}_k - \boldsymbol{\Phi}_{k,k-1}\hat{\hat{\mathbf{X}}}_{k-1}. \tag{78}$$

The fading filtering result is the same as (10).

If the covariance matrix of the estimated state vector $\hat{\mathbf{X}}_{k-1}$ at epoch t_{k-1} is inflated, and the $\hat{\mathbf{X}}_{k-1}$ is treated as a stochastic vector which is like the observational vector, then the corresponding risk function is like (Yang and Gao 2006c)

$$\Omega(k) = \mathbf{V}_k^T\boldsymbol{\Sigma}_k^{-1}\mathbf{V}_k + \frac{1}{\lambda_k}\mathbf{V}_{\hat{\mathbf{X}}_{k-1}}^T\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}^{-1}\mathbf{V}_{\hat{\mathbf{X}}_{k-1}} + \hat{\mathbf{w}}_k^T\boldsymbol{\Sigma}_{\mathbf{w}_k}^{-1}\hat{\mathbf{w}}_k = \min, \tag{79}$$

where λ_k is the fading factor which satisfies $\lambda_k \geq 1$. The corresponding covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}$ follows:

$$\boldsymbol{\Sigma}_{\overline{\mathbf{X}}_k} = \lambda_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}\boldsymbol{\Phi}_{k,k-1}^T + \boldsymbol{\Sigma}_{\mathbf{W}_k}, \tag{80}$$

where $\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}$ is the covariance matrix of the state estimated at epoch t_{k-1}, which results in fading filtering.

The obvious difference between the fading filter and the standard Kalman filter is that the prior state covariance matrix in the fading filter is inflated by λ_k times in order to reduce the contribution of the prior state and strengthen the contribution of the present measurements on the last state estimate.

Another fading filter is based on the following risk function:

$$\Omega(k) = \mathbf{V}_k^T\boldsymbol{\Sigma}_k^{-1}\mathbf{V}_k + \mathbf{V}_{\hat{\mathbf{X}}_{k-1}}^T\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}^{-1}\mathbf{V}_{\hat{\mathbf{X}}_{k-1}} + \frac{1}{\lambda_k}\hat{\mathbf{w}}_k^T\boldsymbol{\Sigma}_{\mathbf{w}_k}^{-1}\hat{\mathbf{w}}_k = \min. \tag{81}$$

Then

$$\boldsymbol{\Sigma}_{\bar{\mathbf{X}}_k} = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}\boldsymbol{\Phi}_{k,k-1}^T + \lambda_k\boldsymbol{\Sigma}_{\mathbf{W}_k}. \tag{82}$$

The two fading filters above are very similar, and both of them are different from the standard Kalman filter with the risk function (8) or

$$\Omega(k) = \mathbf{V}_k^T\boldsymbol{\Sigma}_k^{-1}\mathbf{V}_k + \mathbf{V}_{\hat{\mathbf{X}}_{k-1}}^T\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}^{-1}\mathbf{V}_{\hat{\mathbf{X}}_{k-1}} + \hat{\mathbf{w}}_k^T\boldsymbol{\Sigma}_{\mathbf{w}_k}^{-1}\hat{\mathbf{w}}_k = \min. \tag{83}$$

The key problem of fading filter is to construct a reasonable fading factor. Two kinds of fading factors, based on the optimization theory, are established (Fagin 1964; Fang 1998), one of which is expressed as

$$\lambda_k = \max\{1, \frac{1}{n}\mathrm{tr}(\mathrm{N}_k\mathrm{M}_k^{-1})\}, \tag{84}$$

where tr[·] denotes the trace of matrix. $\mathbf{M}_k$ and $\mathbf{N}_k$ are defined as

$$\mathbf{M}_k = \mathbf{A}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}\boldsymbol{\Phi}_{k,k-1}^T\mathbf{A}_k^T, \tag{85}$$

$$\mathbf{N}_k = \boldsymbol{\Sigma}_{\bar{\mathbf{V}}_k} - \mathbf{A}_k\boldsymbol{\Sigma}_{\mathbf{W}_k}\mathbf{A}_k^T - \boldsymbol{\Sigma}_k, \tag{86}$$

where $\boldsymbol{\Sigma}_{\bar{\mathbf{v}}_k}$ is the covariance matrix of the predicted residual vector $\bar{\mathbf{V}}_k$ (Yang et al. 2001; Yang and Xu 2003):

$$\boldsymbol{\Sigma}_{\bar{\mathbf{V}}_k} = \mathrm{E}(\bar{\mathbf{V}}_k\bar{\mathbf{V}}_k^T). \tag{87}$$

Usually $\boldsymbol{\Sigma}_{\bar{\mathbf{v}}_k}$ is calculated by windowing estimation method (Xia et al. 1990), similar to the Sage filtering (Yang and Xu 2003), that is,

$$\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{V}}_k} = \frac{1}{k-1}\sum_{i=1}^{k}\bar{\mathbf{V}}_i\bar{\mathbf{V}}_i^T. \tag{88}$$

One simpler expression of (84) is defined as follows (Sorensen and Sacks 1971):

$$\lambda_k = \max\{1,\ \mathrm{tr}(\mathbf{N}_k/\mathbf{M}_k)\}. \tag{89}$$

Theoretically, the fading factor λ_k above is optimal. Increasing the predicted residual vector $\bar{\mathbf{V}}_{\bar{\mathrm{X}}_k}$ will increase the covariance matrix $\boldsymbol{\Sigma}_{\bar{\mathbf{v}}_k}$ based on (88) and results in an optimal fading factor λ_k.

Formula (88) can be improved as (Fang 1998)

$$\boldsymbol{\Sigma}_{\bar{\mathbf{V}}_k} = \begin{cases} \frac{\lambda_{k-1}\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T}{1+\lambda_{k-1}}, & k > 1, \\ \frac{1}{2}\bar{\mathbf{v}}_\mathbf{0}\bar{\mathbf{v}}_\mathbf{0}^T, & k = 1, \end{cases} \tag{90}$$

where $\bar{\mathbf{v}}_\mathbf{0}$ is the predicted residual vector when $k = 0$.

The improved $\boldsymbol{\Sigma}_{\bar{\mathbf{v}}_k}$ expressed by (90) is more sensitive than (88) in reflecting the kinematic model errors at present epoch, since $\boldsymbol{\Sigma}_{\bar{\mathbf{v}}_k}$ in (90) does not average the historical information, which uses the present information directly.

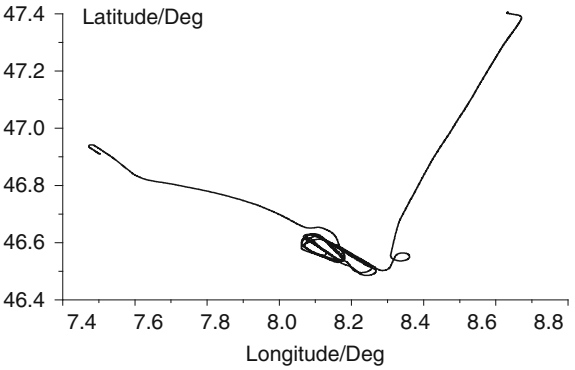


Fig. 2.10 The orbit of the airplane movement

2.6.2 Comparison of Fading Filter and Adaptive Filter

The adaptively robust filter mentioned above is intermediate between fading filter and standard Kalman filter. It does not distinguish the errors of $\hat{\mathbf{X}}_{k-1}$ from the kinematic model errors. It treats the predicted state vector $\overline{\mathbf{X}}_k$ as a whole and adopts the principle that $\alpha_k \mathbf{V}_{\overline{\mathbf{X}}_k}^T \boldsymbol{\Sigma}_{\overline{\mathbf{X}}_k}^{-1} \mathbf{V}_{\overline{\mathbf{X}}_k}$ is a minimum, in which the adaptive factor α_k changes between [0, 1]. When the kinematic model error increases or the vehicle movement is in an unstable state (Fig. 2.10), α_k is smaller than 1 or even equals 0. When the kinematic model error is small enough, α_k equals 1, and the adaptive filtering changes into the standard Kalman filtering.

Analysing the adaptively robust filtering and fading filtering, we find that the primary differences are as follows:

1. The adaptively robust filtering adopts the principle of robust estimation, and it can control the influences of the measurement outliers on the estimated state vector.
2. The adaptive factor α_k acts on the covariance matrix of the predicted state vector $\overline{\mathbf{X}}_k$, while the fading factor acts on the covariance matrix of the previous estimated state vector $\hat{\mathbf{X}}_{k-1}$.
3. In the fading factor, the matrix $\mathbf{N}_k$ expressed by (86) may be indefinite, which usually leads to the failure of the filter. If the adaptive filter is expressed by (14), the adaptive factor can be changed in [0, 1]. If the adaptive filter is expressed by (15) and (16), then the adaptive factor can be changed in [0, 1]. The adaptive factor is determined by observational information and state predicted information, which is capable of adapting in real time. Because the adaptive factor is constructed by the discrepancy between the predicted state vector and the estimated state vector by measurements, it has strong adaptation ability and real time flexibility.

2.6.3 Actual Computation and Analysis

The actual data was collected by using two Trimble 4000SSI receivers mounted on an aircraft. To analyse the roles of the adaptive factors in the adaptive filtering, the highly precise results from double-differenced carrier measurements were used as "true values" to compare with the results from the code measurements. The constant-velocity model of the Kalman filtering was employed. The initial variances for positions, velocities and C/A-code were selected separately as 0.2 m^2, 9×10^{-5} m^2/s^2 and 1.0 m^2, respectively. The spectral density for velocities was chosen to be 0.01 m^2/s^2. The selected kinematic model covariance matrix was the same as that used in Jazwinski (1970) and the following three schemes were carried out:

- Scheme 1: Standard Kalman filtering (SKF)
- Scheme 2: Fading Kalman filtering (FKF)
- Scheme 3: Adaptively robust Kalman filtering (ARKF)

The errors of the X axis of the three schemes relative to the "true values" are plotted in Figs. 2.11, 2.12 and 2.13. Because the errors in the X, Y and Z axes are similar, only the errors of the X axis are given. The comparison of RMS is shown in Table 2.2.

From the calculation results above, we find that

1. The influences of the disturbances during the flight on the standard Kalman filtering are very significant; see Fig. 2.11 and Table 2.2.
2. From Figs. 2.11 and 2.12 and Table 2.2, we find that the fading filtering can control the influences of the kinematic model disturbances on the navigation results to a certain extent, and the results are obviously superior to the standard Kalman filtering.
3. The results of adaptively robust Kalman filtering are slightly superior to the fading filtering. The adaptively robust Kalman filtering can not only control the

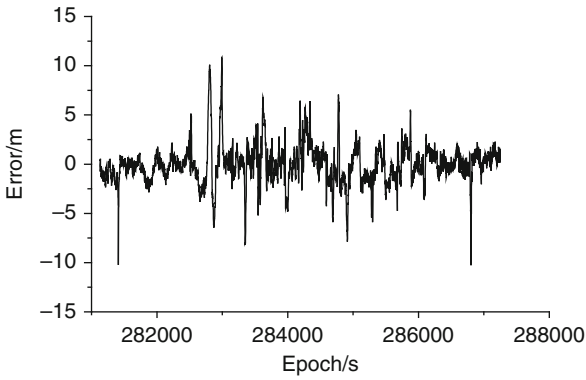


Fig. 2.11 Errors of scheme 1

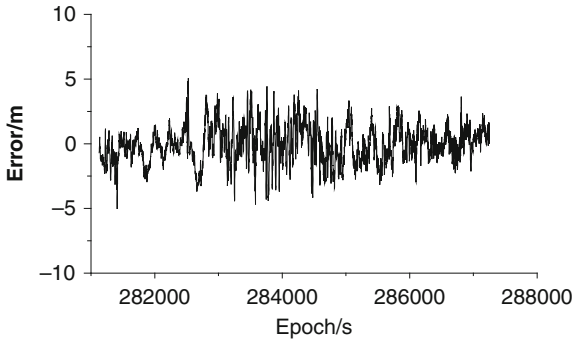


Fig. 2.12 Errors of scheme 2

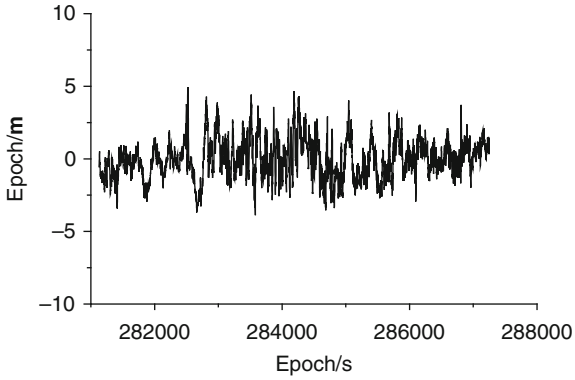


Fig. 2.13 Errors of scheme 3

Table 2.2 Comparison of RMS (unit: m)

		CKF	FKF	ARKF
RMS	X	2.006	1.415	1.331
	Y	1.338	0.759	0.675
	Z	1.936	1.539	1.452

influences of the kinematic model disturbances but also control the influences of the measurement outliers on the navigation results; see Fig. 2.13 and Table 2.2. The results of the adaptively robust filtering are very stable and robust, and the calculation is very flexible.

In conclusion, the fading filtering can control the influences of the kinematic model disturbances on the navigation results to some extent, and its results are obviously superior to standard Kalman filtering. The fading filtering uses the fading factor to inflate the covariance matrix of the estimated state vector at the former

epoch in order to reduce the influences of the state model errors on the new estimated results, but it is difficult to distinguish the model errors from the errors of the former estimated state vector. When the disturbances of the kinematic model are large enough, it is difficult to control their influences by the fading factor λ_k.

2.7 Comparison of Sage Adaptive Filter and Adaptively Robust Filter

In the adaptive Kalman filtering algorithms, use of the Sage–Husa filter (Sage and Husa 1969) is very popular for approaching the variance–covariance matrices by the windowing method (see Jazwinski 1970) and keeping a good consistency between the predicted residuals and the corresponding statistics. A windowing approach of innovation-based adaptive estimation has been studied by Mehra (1970). It makes the covariance matrices of the observation equation and the state errors adapt to the observation information (Mohamed and Schwarz 1999).

The principle of the Sage adaptive filter is

$$\Omega(k) = \mathbf{V}_k^T \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{V}_k + \mathbf{V}_{\overline{\mathbf{X}}_k}^T \hat{\boldsymbol{\Sigma}}_{\overline{\mathbf{X}}_k}^{-1} \mathbf{V}_{\overline{\mathbf{X}}_k} = \min, \tag{91}$$

where $\hat{\boldsymbol{\Sigma}}_k$ and $\hat{\boldsymbol{\Sigma}}_{\overline{\mathbf{X}}_k}$ are estimated by window method.

If the covariance matrix of the current observational errors is computed by the innovation vectors from the previous m epochs, then the adaptive filter is called IAE (innovation-based adaptive estimation) filter. If the covariance matrix is computed by the residual vectors, then the adaptive filter is called RAE (residual-based adaptive estimation) filter. These two adaptively windowing estimations have appeared many times in the literature (Mohamed and Schwarz 1999; Wang et al. 1999; Hu and Ou 1999).

2.7.1 IAE Windowing Method

Suppose that the observational errors are normally distributed. If the width of moving windows is chosen as m, the estimators $\hat{\boldsymbol{\Sigma}}_{\overline{\mathbf{V}}_k}$ of the covariance matrix $\boldsymbol{\Sigma}_{\overline{\mathbf{V}}_k}$ can be given by

$$\hat{\boldsymbol{\Sigma}}_{\overline{\mathbf{V}}_k} = \frac{1}{m} \sum_{j=0}^{m} \bar{\mathbf{V}}_{k-j} \bar{\mathbf{V}}_{k-j}^T. \tag{92}$$

Considering (24) we have the relation of the covariance matrices of the measurement vector, predicted residual vector and predicted state vector as

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{\overline{\mathbf{V}}_k} - \mathbf{A}_k \boldsymbol{\Sigma}_{\overline{\mathbf{A}}_k} \mathbf{A}_k^T. \tag{93}$$

Substituting (92) into (93), we obtain the covariance matrix $\hat{\mathbf{\Sigma}}_k$ of observation information at epoch t_k as

$$\hat{\mathbf{\Sigma}}_k = \hat{\mathbf{\Sigma}}_{\overline{\mathbf{V}}_k} - \mathbf{A}_k \hat{\mathbf{\Sigma}}_{\overline{\mathbf{X}}_k} \mathbf{A}_k^T. \tag{94}$$

2.7.2 RAE Windowing Method

Similar to (92), the covariance matrix of the observational residual vector $\mathbf{V}_k$ can be expressed as

$$\hat{\mathbf{\Sigma}}_{\mathbf{V}_k} = \frac{1}{m} \sum_{j=0}^{m} \mathbf{V}_{k-j} \mathbf{V}_{k-j}^T. \tag{95}$$

From (95), we can estimate the covariance matrix $\mathbf{\Sigma}_k$ of the observational vector at epoch t_k as

$$\hat{\mathbf{\Sigma}}_k = \hat{\mathbf{\Sigma}}_{\mathbf{V}_k} + \mathbf{A}_k \hat{\mathbf{\Sigma}}_{\hat{\mathbf{X}}_k} \mathbf{A}_k^T. \tag{96}$$

In order to estimate adaptively $\hat{\mathbf{\Sigma}}_k$ from (96), $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{X}}_k}$ and residual vector $\mathbf{V}_k$ at epoch t_k are required, while to estimate $\hat{\mathbf{\Sigma}}_{\overline{\mathbf{X}}_k}$ we must first have $\hat{\mathbf{\Sigma}}_k$. Therefore, the covariance matrix of the observational vector at epoch t_k can only be computed using the measurement residuals from the previous m epochs before t_{k-1}:

$$\hat{\mathbf{\Sigma}}_{\mathbf{V}_{k-1}} = \frac{1}{m} \sum_{j=1}^{m+1} \mathbf{V}_{k-j} \mathbf{V}_{k-j}^T. \tag{97}$$

Then (96) can be changed into

$$\hat{\mathbf{\Sigma}}_k = \hat{\mathbf{\Sigma}}_{\mathbf{V}_{k-1}} + \mathbf{A}_{k-1} \mathbf{\Sigma}_{\hat{\mathbf{X}}_{k-1}} \mathbf{A}_{k-1}^T. \tag{98}$$

After having $\hat{\mathbf{\Sigma}}_k$, the weight matrix, $\mathbf{P}_k$, of the observational vector at epoch t_k is computed.

Comparing the IAE estimators (92) and (93) with the RAE estimators (97) and (98), we can make following inferences:

1. The covariance matrix $\hat{\mathbf{\Sigma}}_k$ estimated by IAE includes the errors of predicted state vector $\overline{\mathbf{X}}_k$. The larger the error of $\overline{\mathbf{X}}_k$, the larger the error of $\bar{\mathbf{V}}_k$, which leads to a poor reliability of the covariance matrices $\hat{\mathbf{\Sigma}}_{\bar{\mathbf{V}}_k}$ and $\hat{\mathbf{\Sigma}}_k$.
2. The covariance matrix $\hat{\mathbf{\Sigma}}_k$ estimated by RAE is indeed the covariance matrix $\hat{\mathbf{\Sigma}}_{k-1}$ at epoch t_{k-1}. In order to compute $\hat{\mathbf{\Sigma}}_{\bar{\mathbf{V}}_k}$ and $\hat{\mathbf{\Sigma}}_k$ by using the residual vector $\mathbf{V}_k$, we have to compute $\hat{\mathbf{X}}_k$ at first, while to compute $\hat{\mathbf{X}}_k$ we must have $\hat{\mathbf{\Sigma}}_k$.

Thus we have to use the residuals before the epoch t_{k-1}. The reliability of this prediction depends on the consistency between the measurement accuracy of the current epoch and those of the previous epochs. Otherwise, the representation and reliability of the covariance matrix from this prediction can hardly be ensured.

3. The covariance matrix $\hat{\mathbf{\Sigma}}_k$ estimated by IAE is likely to be negative-definite, that is, the covariance matrix $\hat{\mathbf{\Sigma}}_{\bar{\mathbf{V}}_k}$ is possibly smaller than $\mathbf{A}_k \mathbf{\Sigma}_{\overline{\mathbf{X}}_k} \mathbf{A}_k^T$.
4. In general, the covariance matrix $\hat{\mathbf{\Sigma}}_{\bar{\mathbf{V}}_k}$ computed from (92) is far smaller than $\hat{\mathbf{\Sigma}}_{\bar{\mathbf{V}}_k}$ computed from (90).
5. The computation of $\hat{\mathbf{\Sigma}}_k$ based on IAE or RAE needs the measurement residuals or the innovation vectors from the previous m epochs, which increases the storage load of previous information. In addition, the width of moving window m is difficult to determine.
6. The covariance matrix $\hat{\mathbf{\Sigma}}_k$ computed from (93) or (96) is an average of previous accuracy information, which is almost impossible to use to describe the undulations of the observations at the present epoch. So this kind of adaptive estimation is difficult to use for realizing a real self-adaptation.
7. To estimate the covariance matrix $\hat{\mathbf{\Sigma}}_k$ whether using IAE or RAE requires that observational vectors not only have the same dimension at all epochs but also be the same observation type. Otherwise it is impossible to compute the covariance matrix $\hat{\mathbf{\Sigma}}_k$ using (93) or (96). It is a fatal weakness of IAE or RAE to estimate $\hat{\mathbf{\Sigma}}_k$. In turn, it makes IAE or RAE hard to apply in kinematic positioning or navigation.

2.7.3 The Problems of the Windowing Estimation for Covariance Matrix of Kinematic Model

Let the correction vector of the predicted state vector be given by $\Delta\mathbf{X}_k$; then

$$\Delta\mathbf{X}_k = \hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k. \tag{99}$$

It is easy to deduce that

$$\mathbf{\Sigma}_{\mathbf{W}_k} = \mathbf{\Sigma}_{\Delta\mathbf{X}_k} + \mathbf{\Sigma}_{\hat{\mathbf{X}}_k} - \mathbf{\Phi}_{k,k-1} \mathbf{\Sigma}_{\hat{\mathbf{X}}_{k-1}} \mathbf{\Phi}_{k,k-1}^T. \tag{100}$$

It should be noted that $E(\Delta\mathbf{X}_k) = 0$; then the covariance matrix of $\Delta\mathbf{X}_k$ can be expressed as

$$\hat{\mathbf{\Sigma}}_{\Delta\mathbf{X}} = \frac{1}{m} \sum_{j=0}^{m} \Delta\mathbf{X}_{k-j} \Delta\mathbf{X}_{k-j}^T. \tag{101}$$

The estimate $\hat{\mathbf{\Sigma}}_{\mathbf{W}_k}$ of $\mathbf{\Sigma}_{\mathbf{W}_k}$ can be obtained as

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k} = \hat{\boldsymbol{\Sigma}}_{\Delta\mathbf{X}_k} + \boldsymbol{\Sigma}_{\hat{\mathbf{X}}_k} - \boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_{k-1}}\boldsymbol{\Phi}_{k,k-1}^T. \tag{102}$$

The estimation of $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$ with (102) encounters the following problems: (1) the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$ includes the covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_k}$ of the state parameter estimates at epoch t_k; however, the computation of $\boldsymbol{\Sigma}_{\hat{\mathbf{X}}_k}$ needs $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$; (2) the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$ includes $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\Delta}\mathbf{X}_k}$, which is computed by using the $\Delta\mathbf{X}_j = \hat{\mathbf{X}}_j - \overline{\mathbf{X}}_j$ from m epochs and involves $\hat{\mathbf{X}}_k - \overline{\mathbf{X}}_k$ at epoch t_k which also needs $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$; (3) even if we can estimate $\hat{\boldsymbol{\Sigma}}_{\Delta\mathbf{X}_{k-1}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_{k-1}}$ using the previous $\Delta\mathbf{X}_j$ of m epochs from epoch t_{k-1} and take the latter as an approximation of $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$, it is hard to make $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$ adapt to the real kinematic noise level of the motion of the vehicle because the state disturbance at epoch t_k cannot be reliably reflected by the disturbances from the previous m epochs. It is the same case that $\hat{\boldsymbol{\Sigma}}_{\Delta\mathbf{X}_{k-1}}$ estimated with the average of $\Delta\mathbf{X}_j\Delta\mathbf{X}_j^T$ of m epochs cannot reflect the state noise level at epoch t_k, especially when there is a notable sudden change in state.

In order to avoid the problems mentioned above, we directly estimate $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$. Considering (4), (8) and (20), we express $\Delta\mathbf{X}_k$ as

$$\Delta\mathbf{X}_k = -\mathbf{K}_k\bar{\mathbf{V}}_k. \tag{103}$$

Then

$$\boldsymbol{\Sigma}_{\Delta X_k} = \mathbf{K}_k\boldsymbol{\Sigma}_{\bar{\mathbf{V}}_k}\mathbf{K}_k^T. \tag{104}$$

Once the estimate of $\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{V}}_k}$ is obtained with (92), we can estimate $\boldsymbol{\Sigma}_{\Delta\mathbf{X}_k}$ as

$$\hat{\boldsymbol{\Sigma}}_{\Delta\mathbf{X}_k} = \mathbf{K}_k\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{V}}_k}\mathbf{K}_k^T. \tag{105}$$

In the stable state, $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k}$ can be approximately substituted by $\hat{\boldsymbol{\Sigma}}_{\Delta\mathbf{X}_k}$, that is,

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{W}_k} = \mathbf{K}_k\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{V}}_k}\mathbf{K}_k^T. \tag{106}$$

It should be mentioned that there exists another pair of contradictions in the above adaptive filtering process. Increasing the covariance matrix of state noise using the adaptive estimation is equal to decreasing the covariance matrix of observation noise, and vice versa. So if $\boldsymbol{\Sigma}_{\mathbf{W}_k}$ and $\boldsymbol{\Sigma}_k$ are increased or decreased at the same time, a contradiction arises, which sometimes makes an infinite loop and leads to divergence.

2.8 Some Application Examples

Example 1. A flight experiment is chosen as an example. The data sets were collected by using Trimble 4000SSE on a flight (Yang et al. 2001a, b). The available

measurements are C/A-code, P2-code pseudoranges, L1 and L2 carrier phases and Doppler measurements with 1s data rate. The rover receiver was mounted on an aircraft, and the reference receiver was fixed at a site about 1 km from the initial aircraft location. After about 10 min static tracking, the aircraft took off, and the flight time was about 90 min. The flight states have two notable sudden changes, one is close to epoch 1,000 since the plane takes off and the other is between epoch 3,000 and 4,000 since the flight turns round.

The double-differenced C/A-code and P2-code measurements are employed in the test performance. The constant-velocity model of Kalman filter is employed. An initial variance of 0.2 m^2 for positions, of $9 \times 10^{-6} m^2 s^{-2}$ for velocities, of 1 m^2 for code measurements, and with spectral density of 0.2 $m^2 s^{-3}$ for velocities are selected. The dynamic model covariance matrix is chosen as (Schwarz et al. 1989)

$$\Gamma_t \Sigma_{W_t} \Gamma_t^T = \begin{bmatrix} \frac{1}{3} Q_2 \Delta t^3 & \frac{1}{2} Q_2 \Delta t^2 \\ \frac{1}{2} Q_2 \Delta t^2 & Q_2 \Delta t \end{bmatrix},$$

where Q_2 denotes the spectral density for velocities and Δt denotes a sampling time interval.

The highly precise results from double-differenced carrier measurements are used only as "true values" for comparing with the results from the code measurements, in which the ambiguities are resolved on the fly using LAMBDA method (Teunissen et al. 1997). The following two schemes are performed:

- Scheme 1: Standard Kalman filtering, i.e. $\alpha_k = 1$ and $\bar{\mathbf{P}}_k = \mathbf{P}_k$; see Fig. 2.14
- Scheme 2: Adaptive Kalman filtering, in which the adaptive factor α_k is determined by (73) and $\bar{\mathbf{P}}_k = \mathbf{P}_k$; see Fig. 2.15

Figures 2.14 and 2.15 show that the two unstable states of the flight are obviously reflected in the results of the standard Kalman filtering (Fig. 2.14) and the adaptively robust filter does resist the influences of the dynamic model errors (Fig. 2.15).

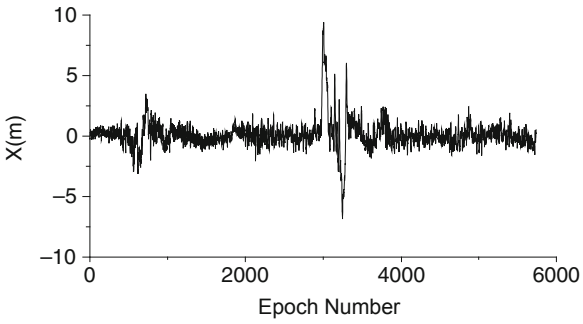


Fig. 2.14 Standard Kalman filtering

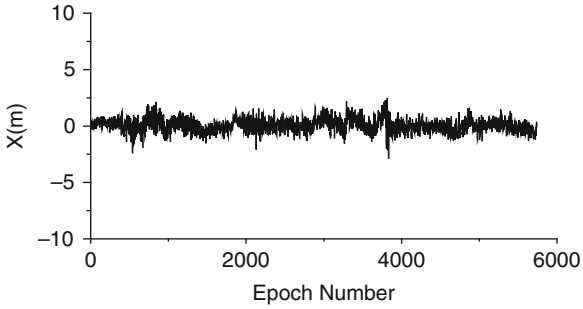


Fig. 2.15 Adaptively robust filtering

(a) **(b)**

Fig. 2.16 Trajectories by differential positioning (*right*), adaptively robust filtering (*middle*) and receiver navigation (*left*)

Example 2. Road information updating by GPS: A GPS receiver is mounted on a vehicle, and a referenced GPS is fixed on a known station (Yang et al. 2004). The pseudorange measurements are employed in the test. Two schemes are performed:

- Scheme 1: Differential GPS positioning
- Scheme 2: Adaptively robust filtering

The two kinds of results and the navigation trajectory results are shown in an image map with the scale 1/50,000; see Fig. 2.16.

Figure 2.16a, b explicitly shows that the receiver navigation results and the differential positioning results have significant systematic errors. If the differential measurements number is less than the number of states, the differential positioning method will not give any position result or give an outlying result. The adaptively robust filtering usually gives reasonable results.

Theoretical analyses and many applications have illustrated that the adaptively robust filter with the corresponding adaptive factors and learning statistics can not

only resist the influences of the measurement outliers but also have a strong ability to control the influences of the state disturbances. It is flexible in computation because the adaptively robust filter is very similar to the standard Kalman filter.

The adaptively robust filter can be applied in some other fields, for example, in crustal deformation, in which the adaptive factor can be used with the geophysical deformation model information, and the robust equivalent weights can be employed with the repeated measurements (e.g. GPS).

References

Cannon, M.E., Schwarz, K.P. and Wong, R.V.C. (1986) Kinematic positioning with GPS: an analysis of road test. Proceedings of the 4th International Geodetic Symposium on Satellite Positioning, Austin, TX, 28 April–2 May

Cui, X. and Yang, Y. (2006) Adaptively robust filtering with classified adaptive factor. Prog. Nat. Sci., 16(8), 846–851

Deng, Z. (2003) Self-Tuning Filtering Theory with Applications—Modern Time Series Analysis Method. Press of Harbin Institute of Technology, Harbin

Fagin, S.L. (1964) Recursive linear regression theory, optimal filter theory and error analysis of optimal system. IEEE Conv. Rec., 12, 216–240

Fang, J. (1998) Study and Application of the Optional Estimation Theories and Method in Integrated Navigation System. Dongnan University, Nanjing (in Chinese)

Hu, G. and Ou, J. (1999) The improved method of adaptive Kalman filtering for GPS high kinematic positioning. Acta. Geodaetica. Et. Cartographica. Sinica., 28(4), 290–294

Huber, P.J. (1964) Robust estimation of a location parameter. Ann. Math. Statist., 35, 73–101

Huber, P.J. (1981) Robust Statistics. Wiley, New York

Jazwinski, A.H. (1970) Stochastic Processes and Filtering Theory. Mathematics in Science and Engineering. Academic Press, New York

Koch, K.R. (2000) Einführung in die Bayes-Statistik. Springer, Heidelberg

Koch, K.R. and Kusche, J. (2002) Regularization of geopotential determination from satellite data by variance components. J. Geod., 76(5), 259–268

Koch, K.R. and Yang, Y. (1998) Robust Kalman filter for rank deficient observation model. J. Geod., 72(8), 436–441

Masreliez, C.J. and Martin, R.D. (1977) Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Autom. Contr., AC-22, 361–371

Mehra, R.K. (1970) On the identification of variances and adaptive Kalman filtering [J]. IEEE Trans. Automat. Contr., 15(2), 175–184

Mohamed, A.H. and Schwarz, K.P. (1999) Adaptive Kalman filtering for INS/GPS. J. Geod., 73, 193–203

Ou, J., Chai, Y. and Yuan, Y. (2004) Adaptive filtering for kinematic positioning by selection of the parameter weights. In: Zhu, Y. and Sun, H. (eds) Progress in Geodesy and Geodynamics. Hubei Science & Technology Press, Hubei, 816–823 (in Chinese)

Panozzo, T., Born, G.H. and Bernelli-Zazzera, F. (2004) Kalman filtering of accelerometric data for aerobraking navigation. Presented in the 6th Conference of Dynamic and Control of Systems and Structures in Space, Riomoggiore, July 2004

Ren, C., Ou, J. and Yuan, Y. (2005) Application of adaptive filtering by selecting the parameter weight factor in precise kinematic GPS positioning. Prog. Nat. Sci., 15(1), 41–46

Sage, A.P. and Husa, G.W. (1969) Adaptive filtering with unknown prior statistics. In Proceeding of Joint Automatic Control Conference, Boulder Colorado,1969, pp. 760–769

Salzmann, M. (1995) Real-time adaptation for model errors in dynamic systems. Bull. Geod., 69, 81–91

Schaffrin, B. (1991) Generating robustified Kalman filters for the integration of GPS and INS. Technical Report, No. 15, Institute of Geodesy, University of Stuttgart, Stuttgart

Schwarz, K.P., Cannon, M.E. and Wong, R.V.C. (1989) A comparison of GPS kinematic models for determination of position and velocity along a trajectory. Manuscripta Geodaetica, 14(6), 345–353

Sorenson, H.W. and Sacks, J.E. (1971) Recursive fading memory filtering. Inform. Sci., 3, 101–119

Teunissen, P.J.G. (1990) Quality control in integrated navigation systems. IEEE Aerosp. Electron. Syst. Mag., 5(7), 35–41

Teunissen, P.J.G., de Jonge, P.J. and Tiberius, C.C.J.M. (1997) Performance of the LAMBDA method for fast GPS ambiguity resolution. Navigation, J. Inst. Nav., 44(3), 373–383

Wang, J., Stewart, M.P. and Tsakiri, M. (2000) Adaptive Kalman filtering for integration of GPS with GLONASS and INS. In: Schwarz, K.P. (ed) Geodesy Beyond 2000. Springer, Heidelberg, 325–330

Wang, Y.J. and Kubik, K.K. (1993) Robust Kalman filter and its geodetic applications [J]. Manuscripta Geodaetica, 18(6), 349–354

Xia, Q., Sun, Y. and Zhou, C. (1990) An optimal adaptive algorithm for fading Kalman filter and its application. Acta Automat. Sin., 16(3), 210–216 (in Chinese)

Xu, G. (2007) GPS Theory, Algorithms and Applications, 2nd Ed. Springer, Berlin

Xu, P. and Rummel, R. (1994) A simulation study of smoothness methods in recovery of regional gravity fields. Geophys. J. Int., 117, 472–486

Xu, T. and Yang, Y. (2000) Modified sage adaptive filtering. Sci. Surv. Mapp., 25, 22–24

Yang, Y. (1991) Robust Bayesian estimation. Bull. Geodesique, 65(3), 145–150

Yang, Y. (1993) The Theory and Application of Robust Estimation. Bayi Publication House, Beijing (in Chinese)

Yang, Y. (1994) Robust estimation for dependent observations. Mans. Geod., 19, 10–17

Yang, Y. (1997) Robust Kalman filter for dynamical system. J. Zhengzhou Inst. Surv. Mapp., 14(2), 79–84 (in Chinese)

Yang, Y. (1999) Theory and Comparisons of Various Robust Filter Models. Collection Papers of Geodesy. Surveying and Mapping Press, Beijing, pp. 51–57 (in Chinese)

Yang, Y., Cheng, M.K., Shum, C.K. and Tapley, B.D. (1999) Robust estimation of systematic errors of satellite laser range. J. Geod., 73, 345–349

Yang, Y. and Cui, X. (2008) Adaptively robust filter with multi adaptive factors. Surv. Rev., 40(309), 260–270

Yang, Y. and Gao, W. (2005) Influence comparison of adaptive factors on navigation results. J. Nav., 58, 471–478

Yang, Y. and Gao, W. (2006a) An optimal adaptive Kalman filter. J. Geod., 80, 177–183

Yang, Y. and Gao, W. (2006b) A new learning statistic for adaptive filter based on predicted residuals. Prog. Nat. Sci., 16(8), 833–837

Yang, Y. and Gao, W. (2006c) Comparison of Two Fading Filters and Adaptively Robust Filter. Geomatics and Inf. Sci. Wuhan Univ., 31(11), 980–982

Yang, Y., He, H. and Xu, G. (2001a) Adaptively robust filtering for kinematic geodetic positioning. J. Geod., 75(2/3), 109–116

Yang, Y., Song, L. and Xu, T. (2002a) Robust estimator for correlated observations based on bifactor equivalent weights. J. Geod., 76(6–7), 353–358

Yang, Y., Song, L. and Xu, T. (2002b) Robust parameter estimation for correlated geodetic observations. Selected Papers for English Edition, Acta Geodeatica Cartography Sinica, pp. 18–24

Yang, Y., Tang, Y., Li, Q. and Zou, Y. (2004) Experiments of adaptive filters for kinematic GPS positioning applied in road information updating in GIS. International Symposium on Inertial Navigation Technology and Intelligent Traffic, Nanjing, October 15–17.

Yang, Y. and Wen, Y. (2003) Synthetically adaptive robust filtering for satellite orbit determination. Sci. China (series D), 32(11), 1112–1119

Yang, Y. and Xu, T. (2003) An adaptive Kalman filter based on Sage windowing weights and variance components. J. Nav., 56(2), 231–240

Yang, Y., Xu, T. and He, H. (2001b) On adaptively kinematic filtering. Selected Papers for English of Acta Geodetica et Cartographica Sinica, pp. 25–32
Zhou, J. (1989) Classical theory of errors and robust estimation. Acta Geodaetica et Cartographica Sinica, 18(2), 115–120 (in Chinese)
Zhou, J., Huang, Y., Yang, Y. and Ou, J. (1997) Robust Least Squares Method. Press of Huazhong Technical University, Wuhan

Chapter 3
Airborne Gravity Field Determination

Rene Forsberg and Arne V. Olesen

Contents

3.1 Introduction

Airborne measurement of gravity has long been a goal for geodesy and geophysics, both to serve geodetic needs (such as geoid determination) and in order to provide efficient and economic mapping of gravity anomalies for geophysical exploration. Although airborne gravimetry has been attempted since the 1960s (LaCoste 1967), it is only in the 1990s, with the development of carrier-phase kinematic GPS methods, that the accuracy has reached a useful level. In later years new gravity acceleration sensors and improved GPS processing methods have resulted in airborne survey accuracies of 1 mGal (10^{-5} m/s^2) or less at a resolution of a few kilometers for several commercial operators (Williams and MacQueen 2001), typically operating in relatively small regions for geophysical exploration and flying during optimal conditions (e.g., at night when turbulence is minimal). For geodesy, however, the need for airborne gravity is more large scale, with airborne gravity providing not

R. Forsberg (✉)
National Space Institute, Technical University of Denmark, Juliane Maries Vej 30, Copenhagen 2100, Denmark
e-mail: rf@space.dtu.dk

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_3,

only an obvious means of gravity field data over large, inaccessible regions, such as polar and mountainous areas, but also an opportunity for seamless surveys across land and marine areas, giving the possibility of accurate geoid models across the coastal zones, useful for both geodetic surveying and oceanographic determination of mean dynamic topography.

The first long-range airborne gravity survey of a continental-scale region was the airborne survey of Greenland in 1991–1992 by the group of the US Naval Research Laboratory, in cooperation with NOAA, NIMA, and the Danish National Survey (Brozena 1992). The survey reached an estimated accuracy level of 4–5 mGal at 12–15 km resolution and was flown at high altitude (4,100 m) and relatively high speed with a large, long-range P3 aircraft. The NRL airborne gravity activities have since been continued and improved in accuracy, with major surveys in 1992–1999 covering large parts of the Arctic Ocean, with an accuracy level of around 2 mGal for the later years (Brozena et al. 1996). Also Russian operators have been performing routine airborne gravity observations, especially over Antarctica and the Arctic, but early surveys such as the work of the Russian Antarctic Expeditions in the 1980s and 1990s (Aleshkova et al. 2000) and the Polar Marine Geological Research Expeditions (PMGRE, Lomonosov) in the Arctic did not apply kinematic GPS methods (positioning was instead based on barometric heights and radionavigation systems), and accuracies were often quite low (5–10 mGal or more). Figure 3.1 shows some of the typical aircrafts used in the 1980–1990s large-aircraft geophysical surveys in the Arctic and Antarctica. Operation of aircraft of this size obviously limited early applications of airborne gravimetry to large countries and organizations.

Implementation of airborne gravity survey systems in smaller aircrafts was done by several research groups in the 1990s, as well as by several commercial companies. Some first applications with modern kinematic GPS positioning were the CASERTZ Antarctic Program (Bell et al. 1992) and the airborne gravity survey of Switzerland (Klingele et al. 1995). In parallel with this several commercial companies such as Carson Geophysics and LCT developed airborne gravity surveys based on smaller aircrafts such as the DHC-6 Twin-Otter or the Cessna Caravan (Fig. 3.2).

Airborne gravity at 1–2 mGal r.m.s. accuracy level in small aircrafts is now fully operational and allows the collection of high-quality airborne gravity data at a typical resolution around 4–6 km. Some private companies, using specially developed

Fig. 3.1 *Left:* Orion P-3 aircraft used in the US NRL airborne gravity program; *right:* IL-76 geophysical laboratory used in the PMGRE Russian polar airborne gravity program. Photos courtesy J. Brozena (NRL) and M. Sorokin (PMGRE)

Fig. 3.2 Twin-Otter (*left*) and Cessna Caravan (*center*) are the most popular aircrafts used for airborne gravity today, due to their excellent performance at low air speeds. *Right picture* shows a typical cabin layout with power conditioner, GPS units, navigation display, and gravimeter in protection box. DTU-Space photos (Jan Mayen Island and Mongolia)

gravity sensors (either modified airborne gravimeter systems, or custom-made, or Russian INS-based systems) now routinely may obtain sub-mGal accuracies for geophysical exploration purposes (Elieff and Ferguson 2008). For high-resolution gravity field mapping, development and application of airborne gravity gradiometers is currently a very active commercial development area for a couple of large companies, as only the use of gravity gradients will allow the very high resolution required for mining applications.

In this chapter we will outline some of the basic principles of airborne gravimetry, with special focus on *geodetic* applications, and give some examples from recent large-scale surveys. For geodesy, the main focus is more on absolute accuracy and long-wavelength stability, since long-wavelength errors in gravity transform to large geoid errors. For geophysical exploration, focus is mainly on the short-wavelength performance, and ultimately making reliable detection and mapping of small, elusive gravity signatures.

With the advent of geodetic satellites such as GRACE and GOCE, the geodetic role of airborne gravity is very much to fill-in the intermediate wavelength bands between satellite gravity (with resolution of hundreds of kilometers) and the resolution of a long-range airborne gravity survey, which could typically be in the 5–10 km range depending on aircraft speed and the needed along-track filtering. For a rough estimate of the variability of the geoid in this wavelength band, the Kaula rule can be used to give an estimate of the omission errors. Based on the Kaula rule, the r.m.s. geoid variation above harmonic degree $n_{\max}$ will be

$$\sigma_N \approx \frac{64}{n_{\max}}[m]. \tag{1}$$

This corresponds for spherical harmonic degree 200 (the expected resolution of GOCE) to around 32 cm r.m.s. and with an airborne survey of 10 km resolution (harmonic degree 2,000) to 3 cm r.m.s. It should be pointed out, however, that the Kaula rule is for the global average behavior of the gravity field and that typical regions for an airborne survey could have a much higher local variability.

3.2 Principles of Airborne Gravimetry

The basic principle of airborne gravimetry is relatively simple. Acceleration measurements in an aircraft will be due to both gravitational attraction $\underline{g}$ and the ficticious forces $\underline{f}$ due to vehicle movement, so an accelerometer triad will measure an apparent force

$$\underline{f} = \underline{g} - \ddot{\underline{r}}. \tag{2}$$

If the position vector $\underline{r}$ of the aircraft is known at all times, e.g., from kinematic GPS, the gravitational acceleration $\underline{g}$ can be found. Equation (2) only holds in an inertial system of reference; in a rotating system, such as the conventional east, north, up (E,N,U) local level system, the equation will be of the form

$$\underline{f} = \underline{g} - \ddot{r} - \underline{\Omega} \times \underline{\Omega} \times \underline{r} + 2\underline{\Omega} \times \dot{\underline{r}} + \dot{\underline{\Omega}} \times \underline{r}. \tag{3}$$

Here the first correction term is the centrifugal force, and the last two terms are the Coriolis and Euler terms. $\underline{\Omega}$ is the rotation vector. Equations (2) and (3) are also the basic equations for inertial navigation, where $\underline{g}$ is assumed known, and the position and velocity instead found from measurements of $\underline{f}$ and $\underline{\Omega}$ measured (by ring laser gyros) or mechanically controlled and computed from position (by a gyrostablilized platform). Alternatively both $\underline{r}$ and $\underline{g}$ can be determined by systems aided by independent position measurements (e.g., by GPS), giving rise to *inertial vector gravimetry.* We will not treat these issues here, as all systems applied for gravimetry in practice until now only are concerned with the *vertical* component ("scalar gravimetry"). A recent review of the principles of inertial navigation and gravimetry can be found in Jekeli (2000).

For the vertical component, with a conventional change of the gravity and acceleration measurement coordinate axis to be positive down, Eq. (3) becomes of the form

$$g_D = a_D - \dot{v}_U - (\dot{\lambda} + 2\omega_{\text{ie}})\cos\varphi \cdot v_E - \dot{\varphi} \cdot v_N, \tag{4}$$

where φ and λ are latitude and longitude, respectively, and ω the sideral earth rate. The Euler term has here been neglected (rotation rate assumed constant); the centrifugal force due to earth's movement is included in normal gravity.

In the commonly used gravimeter systems, the vertical acceleration measurement is done on a gyrostabilized platform. Such a platform would never be perfectly horizontal and have small tilts, which means that the measurement of a_D will be biased. Expressing the gravity g_D as the sum of normal gravity γ and the free-air anomaly Δg,

$$g_D = \gamma + \Delta g = \gamma_0 - \frac{\partial \gamma}{\partial h} H + \Delta g \cong \gamma_0 - 0.3086\,[\text{mGal/m}]H + \Delta g, \tag{5}$$

and noting that gravity is always measured as a *relative* measurement, the fundamental formula for airborne gravity measurement becomes

$$\Delta g = a - h'' - \delta g_{\text{eot}} - \delta g_{\text{tilt}} - y_0 + g_0 - \gamma_0 + 0.3086(h - N), \tag{6}$$

with the notation

a:	the measured acceleration along the vertical
h'':	vertical acceleration derived from GPS
a_0:	airport base reading (zero level of the gravimeter)
g_0:	airport reference gravity value
h:	GPS ellipsoidal height
H:	orthometric height
δg_{tilt}:	gravimeter platform tilt correction (due to the non-verticality of the acceleration sensor)
δg_{eot}:	Eotvos correction (due to the movement of the platform over a curved, rotating earth)
γ_0:	normal gravity at sea level
N:	geoid h*eigh*t

The advantage of formulating Eq. (6) by free-air anomalies is that Δg opposed to g itself only varies relatively weakly with height. It is a common misconception, especially in geophysics, that the free-air anomaly refers to gravity at sea level; free-air anomalies are in the modern view a full three-dimensional quantity, and the variation of the free-air anomaly with altitude is to first order equivalent to the vertical gravity gradient, which can vary strongly in mountainous topography. Since heights in airborne gravimetry can be many kilometers, it is usually not sufficient to use a constant free-air gradient (–0.3086 mGal/m), and the more exact height dependence for normal gravity must be used:

$$\gamma = \gamma_0 + \frac{\partial \gamma}{\partial h} \cdot (H - N) + \frac{\partial^2 \gamma}{\partial h^2} \cdot (H - N)^2, \tag{7}$$

with the gradient terms slightly depending on latitude, with the first term –0.30877 $(1 - 0.00242 \sin^2\varphi)$ and the second for the GRS80 reference field $-0.75 \times 10^{-7} h^2$ [mGal/m^2] for the GRS80 ellipsoid. The difference between the formulas at $h =$ 4 km is more than 1 mGal and therefore significant. When gravity anomalies are to be used for geoid determination, an atmospheric correction must also be applied. This amounts to +0.87 mGal at sea level, but changes with height and is only +0.53 mGal at $h = 4$ km.

The Eotvos correction of the fundamental equation (6) contains a large heading-dependent term and a smaller centrifugal force term and may on a sphere be expressed as

$$\delta g_{\text{eot}} = -2\omega \cos\varphi v_E - (v_N^2 + v_E^2)/R, \tag{8}$$

where ω is the earth rotation rate, R the earth radius, φ the geographic latitude, and v_N and v_E north and east velocity components. For a typical small-aircraft survey speed of 140 knots (e.g., a Twin-Otter), this gives a correction of 1,127 mGal for an E–W flight on the equator. Although large, the correction only changes slowly and may be determined with kinematic GPS with high accuracy.

The off-level correction δg_{tilt} is closely related to the mechanics of the gyrostabilized platform with the gravimeter sensor. Figure 3.3 illustrates the basic principle. Two horizontal accelerometers provide a measurement of apparent horizontal acceleration in two orthogonal directions x and y, but if the platform has tilt angles ε_x and ε_y, the tilt error due to non-vertical measurement of gravity may be shown to be

$$\delta g_{\text{tilt}} = \frac{a^2 - q^2}{2\,g}, \tag{9}$$

with $a = \sqrt{a_x^2 + a_y^2}$ being the measured acceleration and $q = \sqrt{q_E^2 + q_N^2}$ the true horizontal acceleration. The latter can be measured with GPS, and in principle δg_{tilt} can thereby be determined directly; however, the term is non-linear and therefore subject to a serious bias in connection with the necessary filtering. An alternative formulation of the tilt correction problem may be found in Olesen (2002), where a careful modeling of the platform attitude errors gives more linear tilt correction errors.

The careful modeling of the tilt correction is the key for bias-free airborne gravity; in principle most gravimeter sensors, especially of the LCR type, show very low drift (Fig. 3.4). With proper modeling of the correction terms there is therefore in

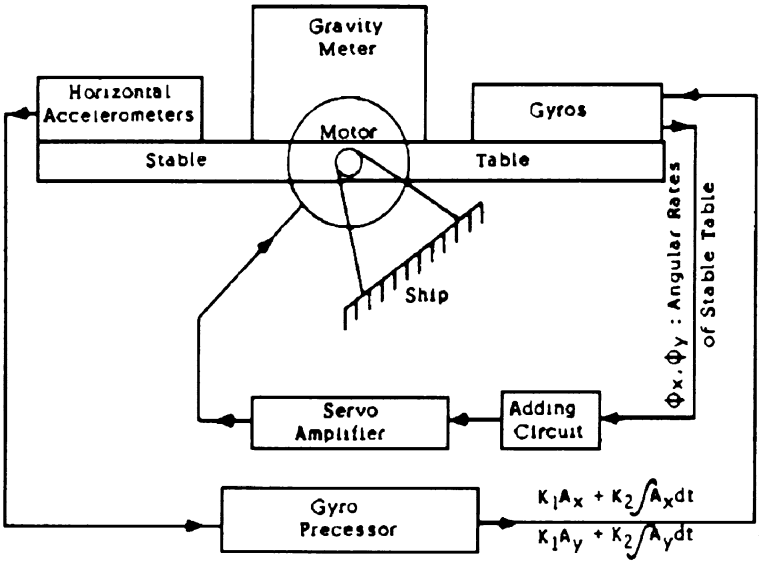


Fig. 3.3 Principle of gyrostabilized platform for the LCR gravimeter. From Valliant (1991)

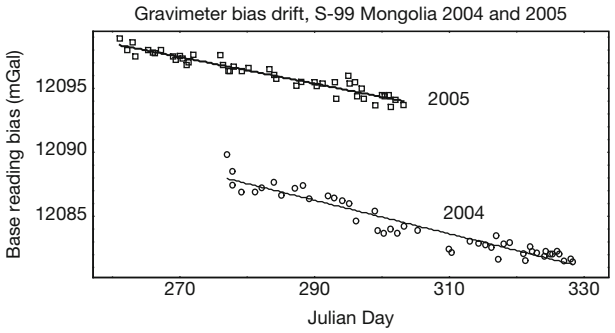


Fig. 3.4 Example of gravimeter drift of an LCR gravimeter (S-99) during 2 month-long airborne gravity campaigns in Mongolia (DTU-Space)

principle no specific need for the widespread use of crossover adjustment of survey results; the sensor by itself is inherently stable at the 1-mGal level. In practice, however, crossover adjustment may eliminate residual, unmodeled errors and often provide the only realistic estimate of survey accuracy since comparisons to ground data are complicated by upward continuation errors.

Other errors include internal measurement and calibration errors, especially the cross-coupling error of lever-beam-type gravimeters such as the LCR S-type marine gravimeter. These corrections are specific to the individual instruments and usually computed in real time, based on in situ calibrations; for details see Valliant (1991). Other sources of errors are in the measurement of acceleration itself; LCR uses a combination of beam drift and spring tension for an overdamped beam, both subject to calibration and unmodeled nonlinearities.

Modern INS-grade accelerometers, such as the ones that were used in the Russian GT-1A gravimeter system (marketed in the west by Canadian MicroGravity) and the Canadian AirGrav system (Sander Geophysics), do not show cross-coupling and related errors. With enhanced INS-style processing and enhanced linearity and sensor performance, such instruments currently provide some of the best results for airborne gravity and may even be used for non-level "draped" survey flights (Studinger et al. 2008). Long-term sensor drift stability can, however, be a major issue for INS-based instruments, with drifts up to many mGal for a flight, as seen in early experiments with strapdown inertial measurement units for gravimetry (Glennie and Schwarz 1999; Glennie et al. 2000).

3.3 Filtering of Airborne Gravity

Inherent to all types of airborne gravity is the need for filtering. GPS can just not measure the double derivative of GPS height h" in Eq. (6) with sufficient accuracy. The terms a and h" are typically two orders of magnitude larger than the wanted quantity Δg; for an example see Fig. 3.5. The magnitude of the ambient vertical

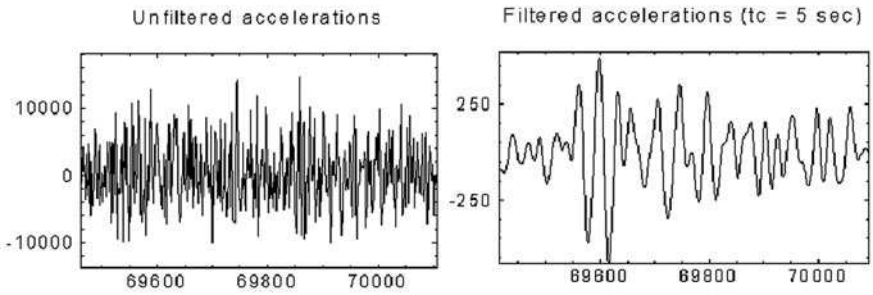


Fig. 3.5 Example of vertical accelerations with and without weak filtering for a smooth flight (Twin-Otter over Disko Bay, Greenland). *x*-axis is time in seconds, *y*-axis acceleration in mGal

accelerations, and inherently the accuracy of airborne gravity, is highly dependent on aircraft type, autopilot performance, and especially the degree of turbulence. Commercial airborne surveys for geophysical exploration are thus often flown at night, where many regions experience more stable air; obviously low-level night flying can be quite dangerous and for large-scale surveys often impossible due to airport restrictions and logistics.

The filtering of airborne gravity is basically a trade-off between resolution and accuracy and must be applied consistently on all terms in Eq. (6). A zero-phase filter is required to avoid shifting of anomalies, and popular types include combinations of forward–backward RC or Butterworth filters or more advanced frequency domain filters. The advantage of the RC and Butterworth filters is that they are very simple sequential filters, readily implemented in just a few lines of code. An example of a typical filter response is shown in Fig. 3.6. For more smooth flight conditions LCR meters are typically processed with filter resolutions around 150–160 sec, corresponding to a 4–5 km resolution (half-wavelength). For comparison, the INS-based gravimeters are typically processed with filters at resolution around 100 sec. This resolution level tends to be the shortest level of filtering used in practice; for shorter

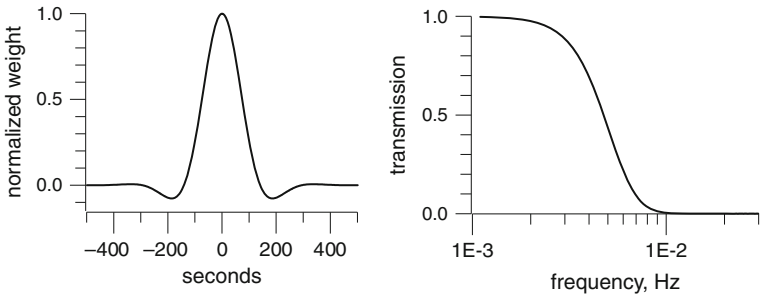


Fig. 3.6 Example of a triple forward/backward Butterworth typical filter for turbulent conditions (200 sec)

filters GPS vertical acceleration errors often rapidly increase beyond 1 mGal.

3.4 Some Results of Large-Scale Government Airborne Surveys

This section outlines results of some selected large-scale airborne surveys covering very different regions of the earth. We use the internal crossover statistics as the indicator for internal accuracy. For the DTU-Space (formerly DNSC) surveys, this crossover analysis is used only for error studies; we find that the airborne gravity performance is sufficiently stable so that no crossover adjustment is necessary in the final results (Olesen et al. 2000). This has large practical advantages, limiting the need for "cross-tie" flights and minimizing the inherent aliasing of errors in the crossover adjustment, an error source of special worry in geoid determination.

The first example is the major airborne surveys of Arctic Ocean regions, carried out in the years 1992–2003 by the NRL, USA, using a large P-3 Orion aircraft (Brozena et al. 1996), and DTU-Space (in cooperation with Canada and Norway), using a Twin-Otter (Forsberg et al. 2001). Figure 3.7 shows the airborne tracks, with all marine flights done at low elevation. Both groups used a LCR model S marine gravimeter, modified by Ultrasys.

Tables 3.1 and 3.2 show the results, based on the statistics that are for the crossovers; therefore the estimate of track noise, assuming uncorrelated errors, will be the r.m.s. divided by $\sqrt{2}$. An accuracy of just below 2 mGal is demonstrated. The increasing accuracy of the NRL results from the large-aircraft surveys in the central Arctic Ocean illustrates the improved performance due to the development of GPS and processing. The error results for both data sets have been confirmed by extensive comparisons to ground data.

Table 3.3 shows some results from surveys which DTU-Space has carried out over other regions of the world (Olesen and Forsberg 2007). The error estimates are based on internal crossover errors and by comparison to GRACE data (GGM03S,

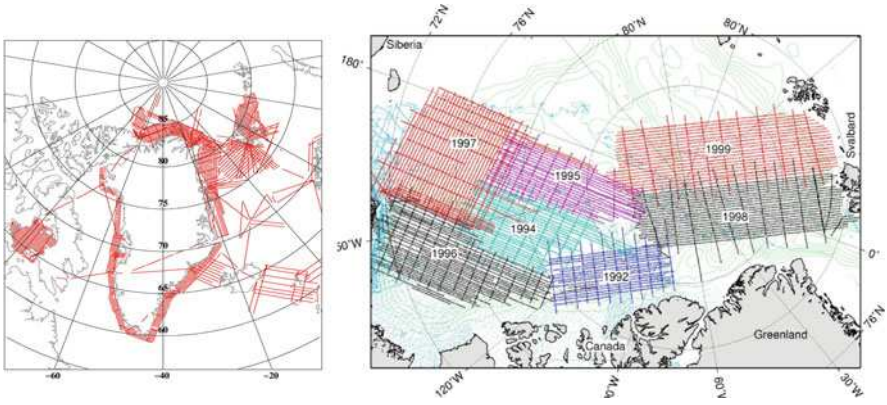


Fig. 3.7 Flight tracks of DTU-Space (*right*) and NRL (*left*) Arctic Ocean and Greenland marine airborne surveys

Table 3.1 Crossover errors for the NRL Arctic surveys (J. Brozena, personal communication)

Data set	Number of x-overs	r.m.s. misfit	r.m.s. error
1998	86	1.8	1.3
1999	74	2.5	1.8
2000	96	2.8	2.0
2001	66	2.6	1.8
2002	101	2.6	1.8
2003	46	2.1	1.5
All years	670	2.5	1.8

Table 3.2 Crossover errors for the DTU-Space Arctic surveys

Data set	r.m.s. misfit	r.m.s. error
1992	4.5	3.2
1994	3.0	2.1
1995	4.1	2.9
1996	2.0	1.4
1997	1.9	1.3
1998	2.6	1.8
1999	2.2	1.6

Table 3.3 Crossover error statistics (mGal) for different large surveys

Data set	No. of crossovers	Line spacing (km)	r.m.s. crossover	Inferred r.m.s. error	Bias to GRACE	Bias to EGM08
Malaysia 2002–2003	1965	5	2.6	1.8	–0.8	0.1
Mongolia 2004–2005	504	18	3.1	2.2	0.5	0.2
Ethiopia 2006–2007	386	18	3.7	2.6	0.6	0.4
Korea 2008–2009	449	10	2.3	1.6	2.4	0.8

Tapley et al. 2007) and combination models (EGM08, Pavlis et al. 2008). The surveys include Malaysia (in cooperation with JUPEM, Malaysia), Mongolia (in cooperation with Mongolian Geodetic Survey, MonMap, and NGA, USA), Ethiopia (in cooperation with Ethiopian Mapping Agency and NGA), and South Korea (in cooperation with University of Seoul). All surveys were challenging large-scale operations over mountainous areas with highly changing weather and turbulence conditions and were flown by either Twin-Otter or Cessna Caravan aircraft. The results show that airborne gravity routinely gives results at 2 mGal accuracy and,

especially, that data are essentially bias-free, making them suitable for geoid determination (some areas like Korea are too small for GRACE to give a reliable bias). It should be noted that the airborne survey results rely solely on gravity ties to the reference airports; no crossover adjustment or other fitting procedures have been used, and no continuation to a common height level has been done (only Korea was flown at near-constant altitude).

3.5 Downward Continuation of Airborne Gravimetry

With airborne gravity data available as along-track filtered gravity values at altitude, most geodetic applications such as geoid determination would require the gravity data to be downward continued, either to the surface of the terrain or the geoid. This is especially important if some surface data already exist; then many geodetic applications would in practice require a joint grid of gravity values to be made. This can be quite a challenge, since the resolution and accuracy of the surface data might be superior to the airborne data; on the other hand, over large regions, surface data might be of varying quality and age and may often be biased. Airborne gravity is well suited to detect such biases and – especially in the marine domain – make older, doubtful gravity data useful again through an adjustment process, where bias-free airborne data can be used to fix bias problems in marine surveys. An example of such application of airborne gravity can be found in Forsberg et al. (2004).

When a dense and quality-controlled distribution of surface data exist, then surface data may alternatively be *upward continued*. The use for this process would typically be to provide independent gravity values at altitude, mainly for quality control of the airborne data, but could also be used directly for (quasi) geoid computations at altitude, with a subsequent downward continuation to zero level. This quasigeoid downward continuation is more stable than downward continuation of gravity.

Upward or downward harmonic continuations of gravity data are applications of classical geodetic boundary value problems and may be evaluated by many different techniques (e.g., pointmass modeling, Fourier methods, or optimal estimation). For an overview see the classical textbooks of Heiskanen and Moritz (1967) or Moritz (1980).

The downward continuation of airborne gravimetry is especially required for the typical large-scale airborne surveys, where flight line elevations may be changing over a wide range dependent on the underlying topography and where already some surface gravity data exist; in this case the method of choice would be least-squares collocation, which is able to use all available data in a consistent way. After doing a downward continuation, either to the geoid level or to the surface of the topography, the airborne data may be handled by standard methods of physical geodesy, for e.g., geoid determination.

For the basic principle of airborne gravimetry downward continuation, let Δg and Δg^* be the gravity anomalies at altitude h and the geoid level ($h = 0$), respectively. By taking the two-dimensional Fourier transform of Δg,

$$F(\Delta g) = \int\int \Delta g(x,y) e^{-i(k_x x + k_y y)} dx dy, \tag{10}$$

the gravity anomalies at the geoid are obtained by

$$F(\Delta g^*) = e^{kh} F(\Delta g), \quad k = \sqrt{k_x^2 + k_y^2}. \tag{11}$$

It is clear that noise in Δg at larger wavenumbers k_x and k_y is strongly amplified by this operation, and direct use of Fourier domain methods is thus mainly done for upward continuation, not downward continuation. For a review of the use of Fourier transformation methods in geodesy, see Schwarz et al. (1990).

To stabilize the downward continuation process by Fourier methods, it may be utilized that the earth's gravity field in general follows the Kaula rule. This empirical spherical harmonic decay law implies for the power spectral density of Δg to proportionality with k^{-4}. Assuming data noise to be white, the optimal Wiener filter downward continuation operator (Nash and Jordan 1978) becomes of the form

$$F(\Delta g*) = F(\Delta g) \frac{e^{kh}}{1 + ck^4}. \tag{12}$$

Here c is a resolution parameter, depending on the ratio of noise to gravity signal covariance (Forsberg and Solheim 1988). In practice the c parameter is chosen on an empirical basis to obtain a suitable smooth downward continued field; some stabilization is implicitly originating gridding of the airborne gravity data onto a regular grid by collocation, since the collocation (grid) estimate by default generates a smooth function, depending on assumed standard deviations of data.

The drawback of the FFT methods (and equivalent integral methods) is that varying altitudes of the airborne data are not easily taken into account, and existing surface data cannot be readily utilized. Therefore least-squares collocation with full spatial covariance models is the method of choice for downward continuation in practice.

In least-squares collocation the gravity anomaly signal s at a ground grid point is estimated from a vector x containing all available surface and airborne data by

$$\hat{s} = C_{sx}[C_{xx} + D]^{-1}. \tag{13}$$

Covariances C_{xx} and C_{sx} are taken from a full, self-consistent spatial covariance model, and D is the (diagonal) noise matrix. Because the gravity field of the earth is known to follow Kaula rule, it is important to select covariance models which have an implied PSD decay in accordance with this. An example of such a self-consistent covariance model on a spherical earth is the Tscherning–Rapp model (Tscherning and Rapp 1974) and for a flat earth the simpler planar logarithmic covariance model (Forsberg 1987). In the latter model, the gravity covariance between gravity anomalies at two altitudes is of the form

$$C(\Delta g^{h_1}, \Delta g^{h_2}) = -\sum_k \alpha_k \log(D_k + \sqrt{s^2 + (D_k + h_1 + h_2)^2}), \tag{14}$$

where α_k are weight factors combining terms relating to depth values ($D_k = D + kT$), with the "free parameters" D and T taking the role analogous to the Bjerhammar sphere depth of spherical collocation and a "compensating depth" attenuation factor. The attenuation of long wavelengths in the model is necessary when a spherical harmonic reference model (EGM) is used. Figure 3.8 shows a typical example of an empirical covariance function and the associated fit by planar logarithmic model. In practice there is a rather large range of D and T parameters which could give a reasonable fit to the data, and thus give more or less the same results of the downward continuation.

For stabilizing the downward continuation, it is essential to use remove–restore methods. This means that the gravity field is split into three terms

$$\Delta g = \Delta g_1 + \Delta g_2 + \Delta g_3, \tag{15}$$

where the first term is due to a spherical harmonic reference field (e.g., EGM08 to a suitable degree), the second term due to the terrain, and the third term due to the residual field. Only the residual terrain-corrected term Δg_3 is then processed in the collocation downward continuation process, with the EGM and terrain terms rigorously computed either at the airborne point locations (for the "remove" step) or on ground (for the "restore").

A suitable terrain correction type to be used is the RTM (residual terrain model) effects. In this method the terrain is removed relative to a reference surface corresponding to the average terrain surface already implicitly present in the EGM term Δg_1, cf. Fig. 3.9. The terrain effects of gravity both aloft and on the ground may be readily computed by either prism integration techniques or Fourier methods, for details see Forsberg (1984) or Schwarz et al. (1990). When applying a terrain reduction to the airborne gravity data, it is important to realize that the terrain effects must be *filtered* with an equivalent filter to the one used for the airborne gravity processing, e.g., with a filter as shown in Fig. 3.6. This means in practice that the results

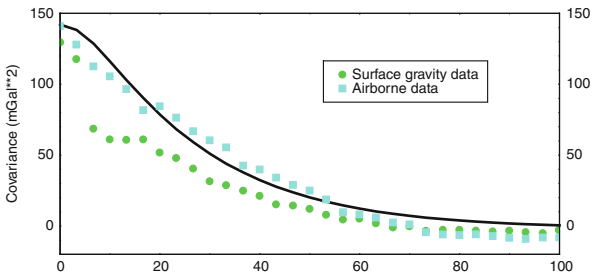


Fig. 3.8 Example of empirical covariance functions for gravity data in Mongolia and the fitted covariance function to airborne data ($D = 10$ km, $T = 49$ km). The longer correlation length of the airborne data is mostly due to the inherent filtering of the airborne data and the attenuation of short-wavelength signals at flight altitude, but also contains signatures of "new" information at the medium wavelengths. Terrain effects and EGM96 removed

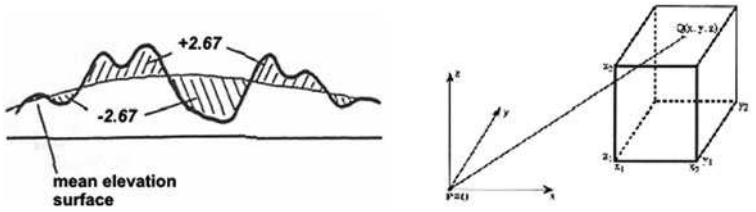


Fig. 3.9 Principle of the RTM terrain reduction. Topography above a smooth reference surface is computationally removed and valleys filled at a reference density of 2.67 g/cm^3 to smooth the data. Integration by rectangular prism elements (*right*) is usually the method of choice

of a prism integration of dense measurement points along a flight track must be followed by an along-track filtering process, where the speed of the aircraft must be taken into account. It can therefore be quite a complicated setup to do the complete downward continuation process.

In the downward continuation process by least-squares collocation, the airborne gravity data represent along-track weighted averages, and the covariance functions used in the least-squares collocation setup must in principle, therefore, also be similarly filtered. A final major constraint in applying least-squares collocation for a large region is that the number of equations to be solved in Eq. (13) may be excessive. A practical solution to this can be to subdivide the area in question into blocks (e.g., $1^\circ \times 1^\circ$ blocks with a 0.5° overlap to neighboring blocks), then do collocation on a blockwise basis, and finally stitch together the computed blocks. This works usually well, in practice, for the downward continuation (but not for direct collocation geoid estimation), as the downward continuation process is essentially a (slight) high-pass filtering operation, whereas the direct geoid computation by collocation is a low-pass filtering operation, much more sensitive to the individual block biases.

3.6 Use of Airborne Gravimetry for Geoid Determination

With a successful downward continuation of airborne gravity data, and possible merging with available surface gravimetry data, the computation of a geoid from the airborne data is equivalent to a conventional geoid computation. Standard methods of physical geodesy can therefore be applied, such as Stokes or Molodensky integration or Fourier methods. These methods all represent variants of the fundamental Stokes integral for obtaining the geoid or quasigeoid, classically for the geoid

$$N = \frac{R}{4\pi\gamma} \int\int \Delta g S(\psi) d\sigma . \tag{16}$$

In practice when removing a spherical harmonic reference gravity field, the used integral should use modified Stokes kernels, e.g., as the modified Wong–Gore kernel

$$S_{\text{mod}}\ (\psi) = S\,(\psi) - \sum_{n=2}^{N_2} \alpha\ (n)\ \frac{2n+1}{n-1} P_n \cos\,(\psi), \tag{17}$$

where

$$\alpha\ (n) = \begin{cases} 1 & \text{for } \ 2 \le n \le N_1 \\ \frac{N_2 - n}{N_2 - N_1} & \text{for } N_1 \le n \le N_2 \\ 0 & \text{for } \quad N_2 \le n \end{cases}, n = 2, \ldots, N\,. \tag{18}$$

The coefficients α prevent the local gravity data to "override" the information inherent in the longer wavelengths in the reference field, which is nowadays determined with very high accuracy from GRACE.

The basic Stokes integral is readily implemented by fast Fourier transforms, e.g., in the "multiband spherical FFT" method, where formula (16) may be virtually exactly evaluated on a sphere, assuming a regular grid in latitude and longitude is available, e.g., from the downward continuation collocation. The basic equation for the geoid determination is of the form

$$N_3 = S_{\text{mod,ref}}(\psi(\varphi, \Delta\lambda)) * [\Delta g_3(\varphi, \lambda) \sin\varphi] = F^{-1}[F(S_{\text{ref}})F(\Delta g \sin\varphi)], \tag{19}$$

where $*$ is a space-domain convolution and F the two-dimensional Fourier transform, for details see Forsberg and Sideris (1993).

3.6.1 Case Story of Mongolian Geoid

In the sequel we will use the Mongolian airborne survey 2004–2005 to illustrate a practical geoid computation. The Danish National Space Center (now DTU-Space) carried out the geoid project in cooperation with the Mongolian Administration of Land Affairs, Geodesy and Carthography (A. Munkhtsetseg), MonMap (M. Saandar), and the University of Bergen, Norway, with support from NGA, USA. The primary purpose of the airborne survey was to provide data for global earth gravity models (EGM08) as well as to provide data for a new geoid of Mongolia, as part of the ongoing GPS modernization of the geodetic infrastructure.

The airborne gravity surveys (Fig. 3.10) were carried out using an Air Greenland Twin-Otter in 2004 and a Cessna Caravan aircraft in 2005. A total of 420 flight hours were flown at a track spacing of 10 nautical miles, with an estimated error of the processed gravity anomaly at the flight level of 2.2 mGal (Table 3.3). Because of the rough topography of Mongolia (Fig. 3.11), flight elevations of individual flight lines varied between 2,100 and 4,800 m, necessitating a formal downward continuation. This was done by least-squares collocation using the planar logarithmic covariance function, as outlined in Section 3.5.

Figure 3.12 shows the processed airborne gravity anomalies at altitude. The correlation of free-air anomalies to topography is evident

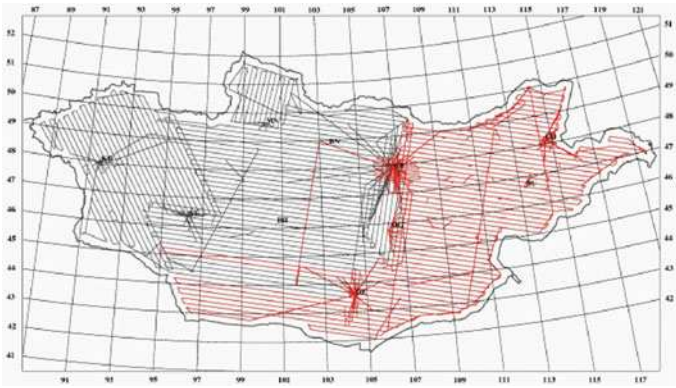


Fig. 3.10 Airborne gravity survey tracks for Mongolia, showing the 2004 (western area) and 2005 (eastern) flights. Flight heights on individual lines range from 2,135 to 4,780 m

In Mongolia older surface data sets were also available, based on earlier Russian and Mongolian gravimetry measurements (Fig. 3.13). The surface data had problems with biases, but were especially useful in supplementing the airborne data along the borders where flights were not possible.

All gravity data were reduced by subtracting a spherical harmonic field to degree 360 (EGM96 blended with GRACE-derived field GGM02S), rigorously computed as a function of latitude, longitude, and height and terrain effects removed by prism integration using the RTM method. The results of this reduction process are shown in Table 3.4. It can be seen that the terrain and EGM96-GRACE reductions very nicely reduce both the data bias and the standard deviations. It is also evident that the surface data are strongly biased, likely due to lack of information on the

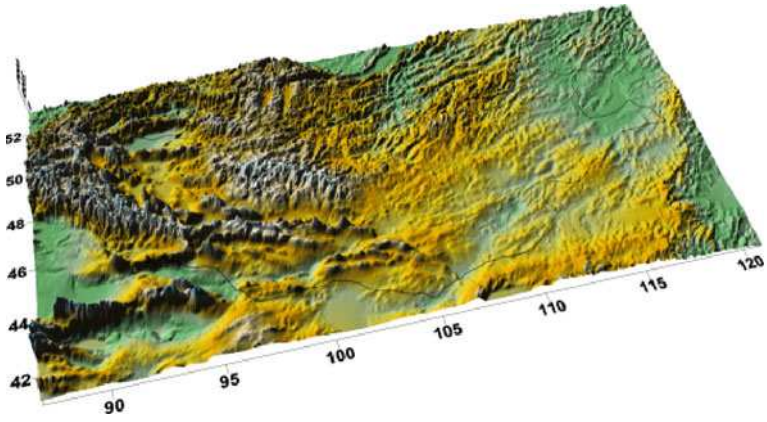


Fig. 3.11 Topography of Mongolia from SRTM data. The elevations range from 4,500 m in western Mongolia to 600 m in the easternmost plains

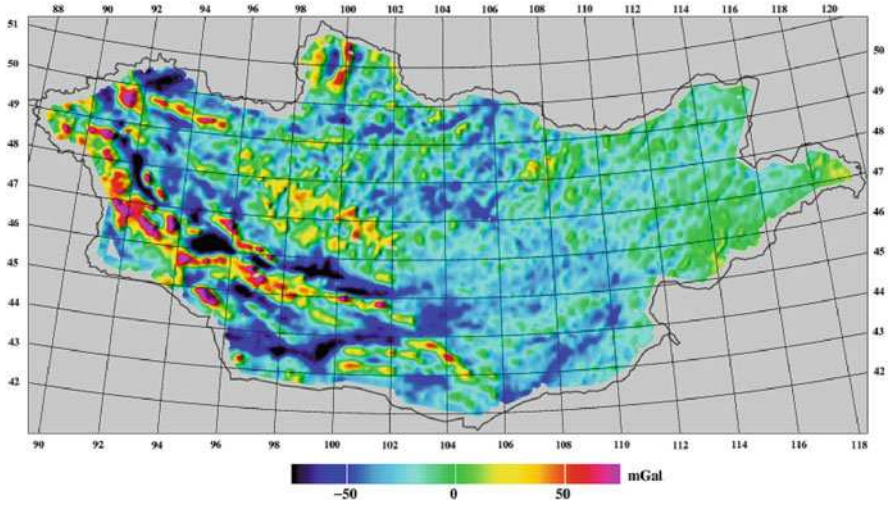


Fig. 3.12 Free-air anomalies at flight altitude from the airborne gravity survey. Color scale is from –75 to 75 mGal

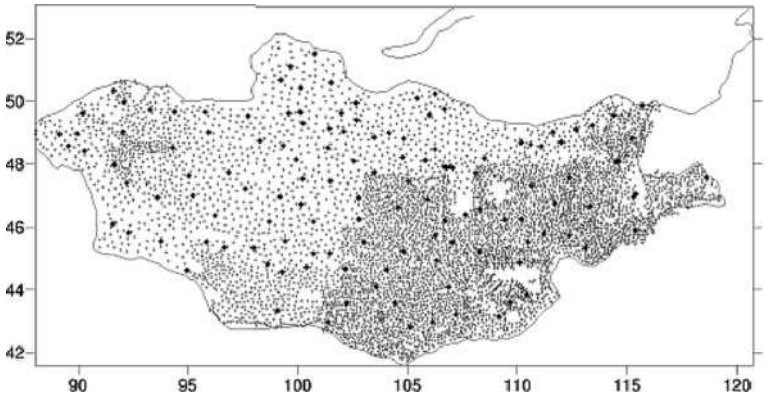


Fig. 3.13 Location of surface gravimetry data in Mongolia. *Thick dots* mark the national reference points and *small dots* data from a GETECH compilation, available only as a 5′ Bouguer gravity grid, and original data locations, from which free-air anomalies were back-interpolated using SRTM data

gravity reference system; the surface data were therefore fitted to the airborne data by subtracting the 12.9 mGal mean offset.

For the downward continuation by collocation, the planar logarithmic model is used, and collocation solutions run blockwise in $1^\circ \times 1^\circ$ blocks expanded with a $0.6^\circ \times 0.8^\circ$ border around the block. Standard errors were assigned to data as follows: airborne gravity 2 mGal, GETECH surface data 3 mGal, and 1 mGal for the national reference station data. Two basic collocation 3′-reduced free-air anomaly

Table 3.4 Statistics of the gravity data reductions (mGal)

Data	Mean	Standard deviation
Airborne gravity data at altitude	17.1	25.9
Airborne minus EGM96/GRACE field	0.2	19.5
Airborne minus EGM96/GRACE and RTM	0.1	11.9
Effect of downward continuation to geoid	0.0	2.5
GETECH interpolated surface gravity data	9.0	19.0
Surface gravity minus EGM96 and RTM	12.9	11.4

grids at the topographic surface were produced: one using all data and one using airborne data only, to study the effect of adding the surface data to the geoid. In addition, the airborne gravity data alone were continued to a common height level grid (3 km), in order to allow a consistent crossover error analysis. The computation was done so that both airborne and surface were downward continued to sea-level anomalies Δg^*, i.e., the downward continued harmonic values, shown in Fig. 3.14. Table 3.4 shows that the effect of continuation was relatively minor on average.

Using the reduced gravity data, gravimetric geoid models were subsequently computed by the multiband spherical FFT method. A modified Wong–Gore Stokes kernel (17) was used, so that only spherical harmonics above degree 40–50 were allowed to modify the underlying GRACE reference field. This process results then in the computation of the reduced quasigeoid at sea level ξ_3^*, to which EGM96 (ζ_1) and RTM terrain effects (ζ_2, Fig. 3.15) are added to give the final gravimetric quasigeoid (Fig. 3.16), after a correction for the difference of quasigeoid between zero height and the terrain level ($\zeta - \zeta^* \approx -\delta g/\gamma . H$). The statistics of the geoid "restore steps" are shown in Table 3.5.

As two geoid solutions were done (one based on airborne gravity data only and one based on both surface and airborne gravity data), the difference of these two

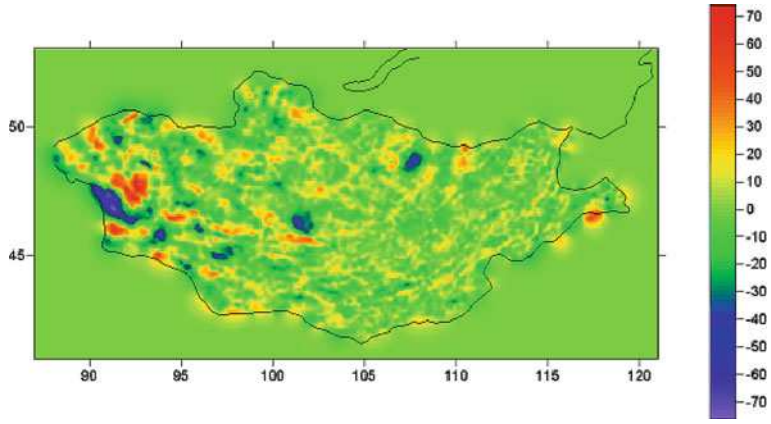


Fig. 3.14 Downward continued airborne gravity data and surface data (mGal)

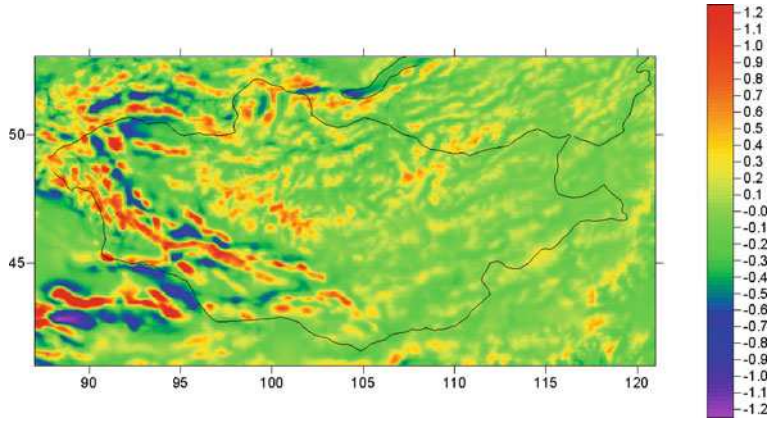


Fig. 3.15 Terrain restore effects (ζ_2) on the geoid, computed from a DEM by FFT (m)

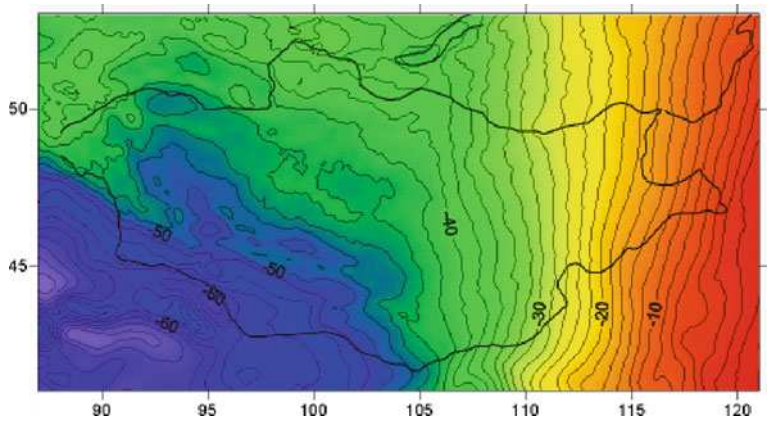


Fig. 3.16 Computed gravimetric quasigeoid of Mongolia from airborne and surface data

solutions will show the changes of adding the (bias fitted) surface data. This is shown in Fig. 3.17. It is seen that the differences are restricted to the border zones, where airborne data were not available, and that the surface and airborne data are otherwise consistent.

For an independent check of the geoid quality, GPS-leveling data points may be used. In Mongolia a set of 58 GPS points on first-order leveling benchmarks were available, giving independent quasigeoid values $\zeta = h^{\mathrm{GPS}} - H_{\mathrm{normal}}$. The comparison statistics showed mean = 1.14 m and standard deviation = 0.20 m. These values mainly reflect the datum offset of the Kronstadt (Baltic) leveling datum of Mongolia, as well as known loop closing errors in the Mongolian first-order leveling network. The GPS-leveling data are therefore not useful for airborne geoid validation.

Table 3.5 Statistics of the geoid restore effects (m)

Geoid grid	Mean	Standard deviation
Fourier transformation of reduced data	0.00	0.27
Terrain effects on geoid	0.01	0.28
Final quasigeoid	–35.94	16.98

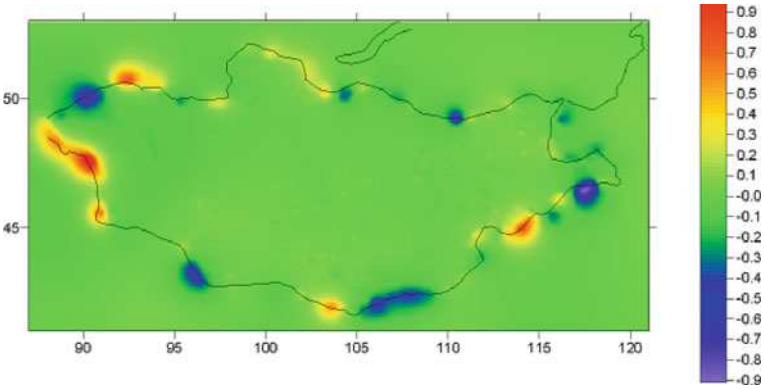


Fig. 3.17 Difference between geoid from all data and geoid from airborne data alone (m)

3.7 Conclusions and Outlook

Airborne gravity is a useful geodetic tool and is a cost-effective way to cover large areas with medium-wavelength gravity data, especially for the purpose of geoid determination, and to map major data gaps in the global terrestrial gravity database, notably in tropical and mountainous regions, the coastal zone, and the polar regions (especially Antarctica). With airborne gravity resolution typically in the range of 3–5 km and upward, the data provide a very useful augmentation of satellite gravity data (GRACE and GOCE) in the spherical harmonic mid-wavelength bands (degrees 90–2,160 and higher) and is a requirement for getting anywhere close to the centimeter-geoid, unless dense surface gravity data are available.

The operational procedures and processing of the airborne gravity data are still a major factor in determining gravity accuracy, and great care must be taken in, e.g., modeling of platform tilt errors to obtain the near-bias-free results indicated in some of the results shown in the present chapter. Some of these processing and operational aspects are closely related to the commonly used LCR platform system. For other systems with more advanced inertial-grade airborne gravity sensors and more rigorous state-space error modeling, e.g., by Kalman filtering and optimal smoothing, more robust and accurate results could be obtained. Such systems could hopefully with time become more widespread and available and within economic bounds of geodetic use.

Ongoing developments of vector gravimetry, taking place in both the commercial and science domains (e.g., Jekeli and Kwon 1999), could potentially make geoid determination more accurate. This is especially true in border regions, where flights close to or beyond a national border are frequently impossible. Unpublished best commercial data on test measurements of horizontal gravity anomalies (i.e., deflections of the vertical) indicate potential accuracies at the 1–2 mGal level (corresponding to 0.2"–0.4" for the deflections). Such accuracies would be most useful for border and line geoid determination, as geoid heights determined from deflections only require simple along-track line integrals of deflections and do not need area integrals like Stokes integral.

References

Aleshkova, N.D., Golynsky, A.V., Kurinin, R.G. and Mandrikov, V.S. (2000) Gravity mapping in the southern Weddell Sea region. Polarforschung, 67, 163–177

Bell, R.E., Coakley, B.J., Blankenship, D.D., Hodge, S.M., Brozena, J.M. and Jarvis, J. (1992) Airborne gravity from a light aircraft: CASERTZ 1990–91. In: Yoshida, Y. (ed) Recent Progress in Antarctic Earth Science, Terrapub, Tokyo, pp. 571–577

Brozena, J.M. (1992) The Greenland aerogeophysics project: airborne gravity, topographic and magnetic mapping of an entire continent. In: Colombo, O. (ed) From Mars to Greenland, Proceedings of the IAG Symposia 110. Springer, New York, pp. 203–214

Brozena, J.M., Peters, M.F. and Salman, R. (1996) Arctic airborne gravity measurement program. In: Segawa, J., Fujimoto, H. and Okubu, S. (eds) Gravity, Geoid and Marine Geodesy, Proceedings of the IAG Symposia 117. Springer, Heidelberg, pp. 131–138

Elieff, S. and Ferguson, S. (2008) Establishing the "air truth" from 10 years of airborne gravimeter data. First Break, 26, 73–77, November 2008

Forsberg, R. (1984) A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modelling. Reports of the Department of Geodetic Science and Surveying, No. 355, Ohio State University, Columbus

Forsberg, R. (1987) A new covariance model for inertial gravimetry and gradiometry. J. Geophys. Res., 92, 1305–1310

Forsberg, R., Olesen, A., Vest, A., Solheim, D., Hipkin, R., Omang, O. and Knudsen, P. (2004) Proceedings of the 2nd International GOCE User Workshop "GOCE, the Geoid and Oceanography", ESA-ESRIN, Frascati, 2004, ESA SP-569

Forsberg, R., Olesen, A.V., Keller, K., Møller, M., Arne Gidskehaug and Solheim, D. (2001) Airborne gravity and geoid surveys in the Arctic and Baltic seas. Proceedings of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS-2001), Banff, June 2001, pp. 586–593

Forsberg, R. and Sideris, M.G. (1993) Geoid computations by the multi-band spherical FFT approach. Manuscripta Geodaetica, 18, 82–90

Forsberg, R. and Solheim, D. (1988) Performance of FFT methods in local gravity field modelling. Proceedings, Chapman Conference on Progress in the Determination of the Earths Gravity Field, American Geophysical Union, Ft. Lauderdale, September 1988, pp. 100–103

Glennie, C.L. and Schwarz, K.P. (1999). A comparison and analysis of airborne gravimetry results from two strapdown inertial/DGPS systems. J. Geod., 73, 311–321

Glennie, C.L., Schwarz, K.P., Bruton, A., Forsberg, R., Olesen, A. and Keller, K. (2000) A comparison of stable platform and strapdown airborne gravity. J. Geod., 74(5), 383–389

Heiskanen, W. and Moritz, H. (1967) Physical Geodesy. Freeman Publishing, San Francisco

Jekeli, C. (2000) Inertial navigation systems with geodetic applications. Walter de Gruyter, Berlin, New York, 352 pp

Jekeli, C. and Kwon, J. (1999) Results of airborne vector (3-D) gravimetry. Geophys. Res. Lett., 26(3), 3533–3536

Klingele, E., Halliday, M., Cocard, M. and Kahle, H.-G (1995) Airborne gravimetric survey of Switzerland. Vermessung, Photogrammetrie, Kulturtechnik, 4, 248–253

LaCoste, L.J.B. (1967) Measurement of gravity at sea and in the air. Rev. Geophys., 5, 477–526

Moritz, H. (1980) Advanced Physical Geodesy. Wichmann Verlag, Karlsruhe

Nash, R.A. and Jordan, S.K. (1978) Statistical geodesy – an engineering perspective. Proc. IEEE, 66, 532–550

Olesen, A.V., Forsberg, R., Keller, K. and Gidskehaug, A. (2000) Airborne gravity survey of Lincoln sea and Wandel Sea, North Greenland, Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, 25(1), 25–29

Olesen, A.V. (2002) Improved airborne scalar gravimetry for regional gravity field mapping and geoid determination. Ph.D. dissertation, National Survey and Cadastre of Denmark Technical Report 24, 123 pp

Olesen, A. and Forsberg, R. (2007) Airborne scalar gravimetry for regional gravity field mapping and geoid determination. Proceedings, Gravity Field of the Earth – 1st Meeting of the International Gravity Field Service, Vol. 73, Issue 18. Harita Dergisi, Ankara, 2007, pp. 277–282

Pavlis, N.K., Holmes, S., Kenyon, S. and Factor, J. (2008) An Earth gravitational model to degree 2160: EGM2008. EGU General Assembly 2008, enna, Austria, April 13–18

Schwarz, K.-P., Sideris, M.G. and Forsberg, R. (1990) Use of FFT methods in physical geodesy. Geophys. J. Int., 100, 485–514

Studinger, M., Bell, R. and Friason, N. (2008) Comparison of AIRGrav and GT-1A airborne gravimeters for research applications. Geophysics, 73(6), 151–161

Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F. and Poole, S. (2007) The GGM03 mean Earth gravity model from GRACE. Eos Trans. Am. Geophys. Union, 88(52), Fall Meet. Suppl., Abstract G42A-03

Tscherning, C.C. and Rapp, R.H. (1974) Closed covariance expressions for gravity anomalies, geoid undulations and deflections of the vertical implied by degree-variance models. Report 208, Department of Geodetic Science, Ohio State University, Columbus

Valliant, H.D. (1991) The LaCoste and Romberg Air/Sea Gravity Meter: An Overview. Geophysical Exploration at Sea, 2nd Ed, Vol. 1. CRC Press, Boca Raton, pp. 141–176

Williams, S. and MacQueen, J.D. (2001) Development of a versatile, commercially proven, and cost-effective airborne gravity system. The Leading Edge, 20(6), 651–654, June 2001

Chapter 4
Analytic Orbit Theory

Guochang Xu

Contents

G. Xu (✉)
GFZ German Research Centre for Geosciences, Department 1: Geodesy and Remote Sensing, Telegrafenberg, 14473 Potsdam, Germany
e-mail: xu@gfz-potsdam.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_4,

4.1 Introduction

The desire to understand the orbits of the planets has a history as long as that of mankind. How and why the planets orbit around the Sun are questions in two categories. One focuses on geometry and the other on physics. Johannes Kepler (1571–1630) answered first the how with his laws of planetary motion in Astronomia nova (1609). Isaac Newton (1643–1727) answered both the how and why with his universal gravitation and laws of motion in *Principia Mathematica* (1687).

Johannes Kepler found that, first, the orbits of the planets in our solar system are elliptical, second, the area velocity (area swept by radius vector within unit time) of a planet is a constant, and third, the ratio of the squares of the period with respect to the cube of the radius is a constant. All three Kepler's laws may be derived theoretically with Newton's laws of planetary motion.

The Keplerian orbit describes the satellite (or planet) motion under the attraction of the central force of the Earth (or the Sun). It is obvious that for satellite orbit of the Earth, the Keplerian orbit is the first approximation. Because of the complication of the Earth's gravitational field and the extraterrestrial disturbances, precise analytic orbit theory is very difficult to derive. The first satellite was launched in 1957. The first-order solution of the equation of satellite motion disturbed by geopotential perturbations was given by William Kaula (1926-2000) in *Theory of Satellite Geodesy* (1966).

Recall Kaula's solution to satellite motion under the influence of the geopotential field. The equations of satellite motion are represented in an inertial coordinate system according to Newton's law. However, the geopotential function is represented in the Earth-fixed system. To transform the geopotential function from the Earth-fixed system to the inertial one, a so-called Kaula's function is created, which is extremely complicated and leads to an extremely complicated solution. Some expressions of the solution are implicit. It is very difficult even to try to get the explicit expressions of the $\overline{C}_{20}$ solutions from Kaula's solution.

After Kaula's theory, studies on orbit theory are mostly based on alternative variables and transformations partly due to the singularity problem in the solution. Many scientists devoted themselves to the second-order orbit solution of geopotential disturbances. The complexity of the theory is such that only a few people understand the theory, and the theory, in turn, is rarely applied in practice. Apparently most studies of the orbit theory are focused on the solution of the geopotential disturbances. Therefore, there exists a blank in the literature on the solution of extraterrestrial disturbances.

Numerical orbit determination is developed directly to meet the needs of the satellite missions and to overcome the problem caused by the missing of analysis solutions of the equations of satellite motion. It appears that the numerical algorithms are very robust and are not affected much by the obvious unphysical models and by the singularity caused by the parameterisation of the problem.

The traditional adjustment model of the solar radiation used in numerical orbit determination is investigated and considered not reasonable physically, and a new

adjustment model is proposed in Xu (2004, 2007). Indeed, one of the ways to obtain the solutions of the extraterrestrial disturbances of the satellite motion is found during that investigation. However, it was not realised until 2007. After the solutions of the extraterrestrial disturbances of the equation of satellite motion were found, great efforts were then made to derive the related solutions of geopotential disturbances. Thereafter, alternative solutions of the extraterrestrial disturbances were found by using different means (also approximated potential function and Gaussian disturbed equations in addition to discretisation). To simplify the solutions, the simplified disturbed equations were proposed. To solve the problem of singularity, the singularity-free theory was also developed. Thanks to the great research freedom granted by GFZ, the complete solution of the orbit equation of motion was found. Based on such a theory, the algorithms of orbit determination can be renewed; a deeper insight into the physics of disturbances becomes possible and the way to a variety of new applications and refinements is opened.

To describe a complete theory of the satellite orbit in a condensed way, perturbed equations of satellite motion are discussed first (Sect. 4.2) after an introduction (Sect. 4.1). Then singularity-free and simplified equations are given (Sect. 4.3). The solutions of extraterrestrial disturbances, such as solar radiation pressure, atmospheric drag and the disturbance of the Sun, the Moon and planets, are then given (Sect. 4.4). Solutions of geopotential disturbances are given in Sect. 4.5. Numerical and analytical orbit determinations are dealt with (Sects. 4.6 and 4.7) before summary and discussions (Sect. 4.8).

The purpose of this chapter is to outline the solutions of the equation of satellite motion. For further details, Kaula (1966/2001) and Xu (2008) are recommended.

4.2 Perturbed Equation of Satellite Motion

Satellites are attracted not only by the central force of the Earth, but also by the non-central force of the Earth, the attracting forces of the Sun and the Moon and the drag force of the atmosphere. They are also affected by solar radiation pressure, Earth and ocean tides, general relativity effects and coordinate perturbations. Equations of satellite motion have to be represented by perturbed equations. In this section, after discussions of the perturbed equations of motion, emphasis is given to the attracting forces and the order estimation of the disturbances.

The perturbed equation of satellite motion is described by Newton's second law in an inertial Cartesian coordinate system as

$$m\ddot{\vec{r}} = \vec{f}\,, \tag{1}$$

where $\vec{f}$ is the summated force vector acting on the satellite and $\vec{r}$ is the radius vector of the satellite with mass m. $\ddot{\vec{r}}$ is the acceleration. Equation (1) is a second-order differential equation. For convenience, it can be written as two first-order differential equations as follows:

$$\begin{aligned} \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} &= \dot{\vec{r}} \\ \frac{\mathrm{d}\dot{\vec{r}}}{\mathrm{d}t} &= \frac{1}{m}\vec{f}. \end{aligned} \tag{2}$$

Denoting the state vector of the satellite as

$$\vec{X} = \begin{pmatrix} \vec{r} \\ \dot{\vec{r}} \end{pmatrix}, \tag{3}$$

(2) can be written as

$$\dot{\vec{X}} = \vec{F}, \tag{4}$$

where

$$\vec{F} = \begin{pmatrix} \dot{\vec{r}} \\ \vec{f}/m \end{pmatrix}. \tag{5}$$

Equation (4) is called the state equation of the satellite motion. Integrating (4) from t_0 to t, one has

$$\vec{X}(t) = \vec{X}(t_0) + \int_{t_0}^{t} \vec{F}\,\mathrm{d}t, \tag{6}$$

where $\vec{X}(t)$ is the instantaneous state vector of the satellite, $\vec{X}(t_0)$ is the initial state vector at time t_0 and $\vec{F}$ is a function of the state vector $\vec{X}(t)$ at time t. Denoting the initial state vector as $\vec{X}_0$, the perturbed satellite orbit problem turns out to be a problem of solving the differential state equation under the initial condition as

$$\begin{cases} \dot{\vec{X}}(t) = \vec{F} \\ \vec{X}(t_0) = \vec{X}_0 \end{cases}. \tag{7}$$

4.2.1 Lagrangian Perturbed Equation of Satellite Motion

If the force $\vec{f}$ includes only the conservative forces, then there is a potential function V so that

$$\frac{\vec{f}}{m} = \mathrm{grad}V = \left(\frac{\partial V}{\partial x}\ \frac{\partial V}{\partial y}\ \frac{\partial V}{\partial z}\right) = \left(\frac{\partial V}{\partial r}\ \frac{\partial V}{\partial \varphi}\ \frac{\partial V}{\partial \lambda}\right), \tag{8}$$

where (x,y,z) and (r,φ,λ) are Cartesian coordinates and spherical coordinates, respectively. Denoting R as the disturbance potential and V_0 as the potential of the centred force $\vec{f}_0$,

$$R = V - V_0, \quad \frac{\vec{f} - \vec{f}_0}{m} = \operatorname{grad} R, \tag{9}$$

the perturbed equation of satellite motion (2) in Cartesian coordinates is then

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \dot{x} \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \dot{y} \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= \dot{z} \\
\frac{\mathrm{d}\dot{x}}{\mathrm{d}t} &= -\frac{\mu}{r^3}x + \frac{\partial R}{\partial x} \\
\frac{\mathrm{d}\dot{y}}{\mathrm{d}t} &= -\frac{\mu}{r^3}y + \frac{\partial R}{\partial y} \\
\frac{\mathrm{d}\dot{z}}{\mathrm{d}t} &= -\frac{\mu}{r^3}z + \frac{\partial R}{\partial z}
\end{aligned}, \tag{10}$$

where μ is the gravitational constant of the Earth. The state vector $(\vec{r}, \dot{\vec{r}})$ of the satellite corresponds to an instantaneous Keplerian ellipse (a,e,ω,i,Ω,M) (i.e. semi-major axis a, eccentricity of the ellipse e, argument of perigee ω, inclination angle i, right ascension of ascending node Ω, mean anomaly M). Using the relationships between the two sets of parameters (for detail see Chap. 3 of Xu 2008), the perturbed equation of motion (10) can be transformed into a so-called Lagrangian perturbed equation system (see, e.g., Kaula 1966/2001):

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{1-e^2}{na^2e}\frac{\partial R}{\partial M} - \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial \omega} \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial e} - \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\left(\cos i\frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega}\right) \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i} \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a} - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e}
\end{aligned}. \tag{11}$$

On the basis of this equation system, Kaula derived the first-order perturbed analysis solution (see Kaula 1966/2001). In the case of a small e (e<<1), the orbit is nearly circular, so that the perigee and the related Keplerian elements f and ω are not

defined (this is not to be confused with the force vector $\vec{f}$ and true anomaly f). To overcome this problem, let $u = f + \omega$, and a parameter set of $(a, i, \Omega, \xi, \eta, \lambda)$ can be used to describe the motion of the satellite, where

$$\begin{aligned} \xi &= e\cos\,\omega \\ \eta &= -e\sin\,\omega\,. \\ \lambda &= M + \omega \end{aligned} \tag{12}$$

The related disturbed equations of motion can be derived (see Sect. 4.4.1.1 of Xu 2008); however, the new variables of (12) do not have clear geometric meanings. Another alternative is to use the Hill variables (see, e.g., Cui 1990).

4.2.2 Gaussian Perturbed Equation of Satellite Motion

Considering the non-conservative disturbance forces such as solar radiation and air drag, no potential functions exist for use; therefore, the Lagrangian perturbed equation of motion cannot be directly used in such a case. The equation of motion perturbed by non-conservative disturbance force has to be derived.

Considering any force vector $\vec{f}=(f_x, f_y, f_z)^{\mathrm{T}}$ in ECSF (Earth-Centred Space-Fixed) coordinate system, one has

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = R_3(-\Omega)R_1(-i)R_3(-u)\begin{pmatrix} f_{\mathrm{r}} \\ f_\alpha \\ f_{\mathrm{h}} \end{pmatrix}, \tag{13}$$

where $(f_{\mathrm{r}}, f_\alpha, f_{\mathrm{h}})^T$ is a force vector with three orthogonal components in an orbital plane coordinate system, the first two components are in the orbital plane, f_{r} is the radial force component, f_α is the force component perpendicular to f_{r} and pointed in the direction of satellite motion and f_{h} completes a right-handed system. R_k is the rotational matrix around the axis k (see Sect. 4.2.2 of Xu 2008). For convenience, the force vector may also be represented by tangential, central components in the orbital plane $(f_{\mathrm{t}}, f_{\mathrm{c}})$ as well as f_{h} (see Fig. 4.1). It is obvious that

$$\begin{pmatrix} f_{\mathrm{r}} \\ f_\alpha \\ f_{\mathrm{h}} \end{pmatrix} = R_3(-\beta)\begin{pmatrix} f_{\mathrm{t}} \\ f_{\mathrm{c}} \\ f_{\mathrm{h}} \end{pmatrix}, \tag{14}$$

where

$$\tan\,\beta = r\frac{\mathrm{d}f}{\mathrm{d}r} = \frac{a(1-e^2)}{1+e\,\cos f}\frac{\mathrm{d}f}{\dfrac{a(1-e^2)}{(1+e\,\cos f)^2}e\,\sin f\,\mathrm{d}f} = \frac{1+e\,\cos f}{e\,\sin f} \tag{15}$$

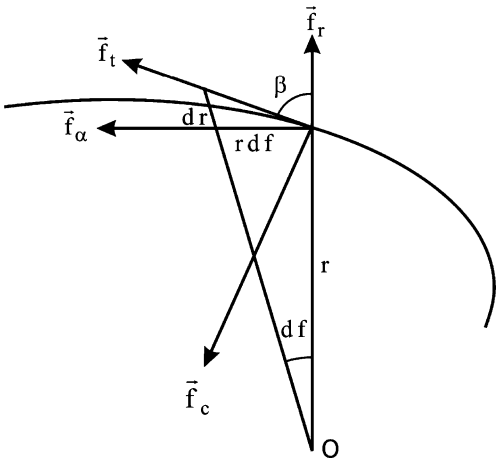


Fig. 4.1 Relation of radial and tangential forces

or

$$\begin{aligned} \sin\,\beta &= \frac{1+e\,\cos f}{\sqrt{1+2e\,\cos f+e^2}} \\ \cos\,\beta &= \frac{e\,\sin f}{\sqrt{1+2e\,\cos f+e^2}} \end{aligned} \quad . \tag{16}$$

To replace the partial derivatives $\partial R/\partial\sigma$ by force components, the relationships between them can be derived (see Sect. 4.4.1.2 of Xu 2008), where σ is a symbol for all Keplerian elements. Putting the relations into Lagrangian perturbed equations of motion (11), the so-called Gaussian perturbed equations of motion are then

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{n\sqrt{1-e^2}}\left[e\,\sin f\cdot f_\mathrm{r}+(1+e\,\cos f)\cdot f_\alpha\right] \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na}\left[\sin f\cdot f_\mathrm{r}+(\cos\,E+\cos f)\cdot f_\alpha\right] \\ \frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{nae}\left[-\cos f\cdot f_\mathrm{r}+\frac{2+e\,\cos f}{1+e\,\cos f}\sin f\cdot f_\alpha\right]-\cos\,i\frac{\mathrm{d}\Omega}{\mathrm{d}t} \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{(1-e\,\cos\,E)\cos\,u}{na\sqrt{1-e^2}}\cdot f_\mathrm{h} \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{(1-e\,\cos E)\sin\,u}{na\sqrt{1-e^2}\,\sin\,i}\cdot f_\mathrm{h} \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= n-\frac{1-e^2}{nae}\left[-\left(\cos f-\frac{2e}{1+e\,\cos f}\right)\cdot f_r+\frac{2+e\,\cos f}{1+e\,\cos f}\sin f\cdot f_\alpha\right] \end{aligned} \quad . \tag{17}$$

The force components of (f_r, f_α, f_h) are used. Using (14), the Gaussian perturbed equations of motion can be represented by a disturbed force vector of (f_t, f_c, f_h).

4.2.3 Keplerian Motion

The simplified satellite orbiting is called Keplerian motion, and the problem is called the two-bodies problem. The satellite is supposed to move in a central force field (i.e. the disturbance potential in (11) or the disturbance force in (17) are zero).

In this case the satellite will orbit in an orbital plane of the mass centre of the Earth. In addition, the moving equation of satellite in the orbital plane is an ellipse, i.e.

$$r = \frac{a(1-e^2)}{1+e\cos f}. \tag{18}$$

The Keplerian motion can be described by six Keplerian elements: inclination angle i, right ascension of ascending node Ω, semi-major axis a, eccentricity of the ellipse e, argument of perigee ω and mean anomaly M. Parameters i and Ω decide the place of the orbital plane, a and e decide the size and shape of the ellipse and ω decides the direction of the ellipse (see Fig. 4.2). Mean anomaly M describes the satellite motion along the ellipse. Three anomalies (eccentric, true and mean anomaly) are equivalent (see Fig. 4.3 and so-called Keplerian equation, Xu 2007).

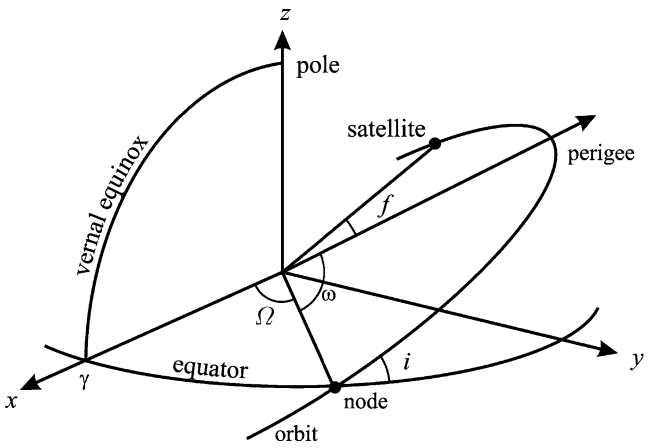


Fig. 4.2 Orbital geometry

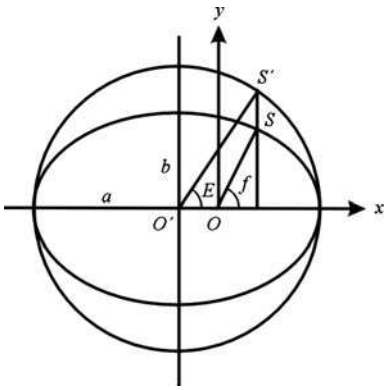


Fig. 4.3 Eccentric and true anomalies of satellite

4.3 Singularity-Free and Simplified Equations

Section 4.2 covered the equations of satellite motion in different forms and the Keplerian orbit. The solutions of equations of motion are given and the singularity-free theory is proposed in Xu (2008). Here the singularity-free theory and simplified equations will be given.

The singularity problem of the solutions of the geopotential disturbances is discussed first. Then the singularity-free disturbed Lagrangian equations of motion are given for three cases, i.e. for the circular orbit, equatorial orbit, circular and equatorial orbit, respectively. If the singularity-free disturbed equations of motion are used, then the derived orbit solutions are singularity-free. Similar discussions are given for the Gaussian equations.

4.3.1 Problem of Singularity of the Solutions

The solutions of Lagrangian and Gaussian equations of (11) and (17) are singular in the cases of $e = 0$ and/or $\sin i = 0$. In other words, the solutions are not valid for the satellite with a circular or an equatorial orbit. An alternative method to overcome the problem of circular orbit has already been discussed in Sect. 4.2.1 by introducing new variables – see (12). The new variables do not have clear geometric meanings and were used to replace the variables (ω, f), which could not be defined in a circular orbit. In the alternative equation of disturbance (cf. Xu 2008), the e factor in the dividend is then eliminated, i.e. the singularity of $e = 0$ disappears. Using another set of variables (a, $h = \sin i \cos \Omega$, $k = -\sin i \sin \Omega$, $\xi = e \cos(\omega + \Omega)$, $\eta = -e \sin(\omega + \Omega)$, $\lambda = M + \omega + \Omega$), both the singularities caused by $e = 0$ and $\sin i = 0$ may disappear. This means that the singularity is not a real problem of the orbits, but a consequence of poor parameterisation of the orbits. Another method to overcome

the singularity problem is the canonical transformation. All these methods overcome the singularity problem on the one hand and pay the price of losing the geometric meanings of the orbital variables on the other.

In the cases of $e = 0$ and/or $\sin i = 0$, the orbits become simpler in practice. However, the equations used to describe a simpler problem become more complicated. This is in conflict with basic scientific philosophy and common sense. A simpler problem should be able to be described in simpler terms.

Looking into the solutions given in Chaps. 5 and 6 of Xu (2008) carefully, it is obvious that the singular problem is not created by the partial derivations of the potential function with respect to the Keplerian variable. In other words, the singularity problem exists from the beginning in the Lagrangian- and Gaussian perturbed equation systems (11) and (17). This may be verified by derivations of (11) (see Kaula 1966/2001).

4.3.2 Disturbed Equations in the Case of Circular Orbit

In the case of a circular orbit, the eccentricity of the ellipse e is a constant of zero; the eccentric anomaly E, true anomaly f and mean anomaly M are identical. Note that in such a case the perigee of the orbit is arbitrary. Then the argument of the perigee ω and the true anomaly f (i.e. mean anomaly M) cannot be separated from each other. However, $\omega + f$, i.e. $\omega + M$ are defined and have the meaning of argument of the perigee plus true anomaly (or mean anomaly) counted from the ascending node of the orbit. For convenience, we write ω and M further separated; in practice, they should be added together. In this special case the orbit is simpler than the general one. The disturbed equations of motion – similar to (11) – can be similarly derived and have accordingly the following simpler forms:

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= -\frac{\cos i}{na^2 \sin i}\frac{\partial R}{\partial i} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= -\frac{1}{na^2 \sin i}\left(\cos i\frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega}\right) \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{1}{na^2 \sin i}\frac{\partial R}{\partial i} \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a}
\end{aligned} \tag{19}$$

4.3.3 Disturbed Equations in the Case of Equatorial Orbit

In the case of an equatorial orbit, sin i, the sine function of inclination angle i, is a constant of zero. Note that in such a case the ascending node is arbitrary. Then the right ascension of the ascending node Ω and the argument of the perigee ω cannot be separated from each other. However, $\Omega + \omega$ are defined and have the meaning of right ascension of the ascending node plus argument of the perigee counted from the vernal equinox. For convenience, we write Ω and ω further separated; in practice, they should be added together. In this special case the orbit is simpler than the general one. Especially, the transformed geopotential function with orbital variable is greatly simplified in such a case. The disturbed equations of motion – similar to (11) – can be similarly derived and have accordingly the following simpler forms:

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{1-e^2}{na^2e}\frac{\partial R}{\partial M} - \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial \omega} \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial e} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a} - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e}
\end{aligned} \quad . \tag{20}$$

4.3.4 Disturbed Equations in the Case of Circular and Equatorial Orbit

In the case of a circular and an equatorial orbit, the eccentricity of the ellipse e is a constant of zero; the eccentric anomaly E, true anomaly f and mean anomaly M are identical; sin i, the sine function of inclination angle i, is a constant of zero. Note that in such a case both the perigee and the ascending node are arbitrary. Then the right ascension of the ascending node Ω and the argument of the perigee ω as well as the true anomaly f (i.e. mean anomaly M) cannot be separated from each other. However, $\Omega + \omega + f$, i.e. $\Omega + \omega + M$ are defined and have the meaning of right ascension of the ascending node plus argument of the perigee plus the true anomaly (or mean anomaly) counted from the vernal equinox. For convenience, we write Ω and ω as well as M further separated; in practice, they should be added together. In this special case, the orbit is the simplest one compared with the others.

The disturbed equations of motion – similar to (11) – in this case can be similarly derived and have accordingly the following simpler forms:

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}\omega}{\mathrm{d}t} &= 0\,. \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a} \end{aligned} \tag{21}$$

4.3.5 Singularity-Free Disturbed Equations of Motion

Define two delta functions as

$$\delta_e = \begin{cases} 1 & \text{if } e \neq 0 \\ e^2 & \text{if } e = 0 \end{cases} \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } \sin i \neq 0 \\ \sin^2 i & \text{if } \sin i = 0 \end{cases}. \tag{22}$$

Then one has the singularity-free disturbed equations of motion

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial M}\delta_e - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial \omega}\delta_e \\ \frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial e}\delta_e - \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i}\delta_i \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\left(\cos i\frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega}\right)\delta_i \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i}\delta_i \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a} - \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial e}\delta_e \end{aligned} \tag{23}$$

Equations (23) are the singularity-free disturbed equations of motion. The solutions derived from these equations are singularity-free. For some reasons, the solutions given in Xu (2008) are mostly derived from (11). To obtain the singularity-free solutions one has to add the two factors of the delta functions (22) to the given solutions and the interested readers may easily attempt these themselves.

4.3.6 Simplified Singularity-Free Disturbed Equations of Motion

The simplified singularity-free disturbed Lagrange equations of motion can be derived and written as

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{na}\frac{\partial R}{\partial M} \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{1-e^2}{2ae}\frac{\mathrm{d}a}{\mathrm{d}t}\delta_e - \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial \omega}\delta_e \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na^2e}\frac{\partial R}{\partial e}\delta_e - \cos i\frac{\mathrm{d}\Omega}{\mathrm{d}t} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\left(\frac{na\cos i}{\sqrt{1-e^2}}\left(\frac{1-e^2}{2}\frac{\mathrm{d}a}{\mathrm{d}t} - ae\frac{\mathrm{d}e}{\mathrm{d}t}\right) - \frac{\partial R}{\partial \Omega}\right)\delta_i \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i}\delta_i \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{2}{na}\frac{\partial R}{\partial a} - \sqrt{1-e^2}\left(\frac{\mathrm{d}\omega}{\mathrm{d}t} + \cos i\frac{\mathrm{d}\Omega}{\mathrm{d}t}\right)
\end{aligned} \tag{24}$$

It is obvious that such equations will lead to a simplified process of solving the problems.

4.3.7 Singularity-Free Gaussian Equations of Motion

Similarly, singularity-free Gaussian perturbed equations of motion are then (cf. (17), (22) and (23)):

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{n\sqrt{1-e^2}}\left[e\sin f\cdot f_{\mathrm{r}} + (1+e\cos f)\cdot f_\alpha\right] \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na}\left[\sin f\cdot f_{\mathrm{r}} + (\cos E + \cos f)\cdot f_\alpha\right] \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{nae}\left[-\cos f\cdot f_{\mathrm{r}} + \frac{2+e\cos f}{1+e\cos f}\sin f\cdot f_\alpha\right]\delta_e - \cos i\frac{\mathrm{d}\Omega}{\mathrm{d}t} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{(1-e\cos E)\cos u}{na\sqrt{1-e^2}}\cdot f_{\mathrm{h}} \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{(1-e\cos E)\sin u}{na\sqrt{1-e^2}\sin i}\delta_i\cdot f_{\mathrm{h}} \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= n - \frac{1-e^2}{nae}\left[-\left(\cos f - \frac{2e}{1+e\cos f}\right)\cdot f_{\mathrm{r}} + \frac{2+e\cos f}{1+e\cos f}\sin f\cdot f_\alpha\right]\delta_e
\end{aligned} \tag{25}$$

The solutions derived from (25) are singularity-free. For some reasons, the solutions given in Xu (2008) are partly derived from (17). To obtain the singularity-free solutions one has to add the two factors of the delta functions (22) into the given solutions and interested readers may easily attempt these themselves. Similar simplified equations can be derived if one wishes.

4.4 Solutions of Extraterrestrial Disturbances

Solutions of the extraterrestrial disturbances of the attracting forces of the Sun, the Moon and planets, the drag force of the atmosphere and solar radiation pressure are given in this section.

4.4.1 Key Notes to the Problems

It is well known that the Keplerian motions of the satellite under the influence of the centre force of the Earth are perfectly and exactly described with mathematical formulas. As soon as it is found by derivation that the satellite is moving in an orbital plane, the equations of motion are re-represented in the plane and the Keplerian motion is then derived. Note that even in the centre force field, it would be nearly impossible to derive the solution without the step of coordinate transformation. This indicates the importance of the selection of the coordinate system. The transformation of the coordinate system is allowed because the frame remained an inertial one after a series of constant rotations.

The use of an alternative coordinate system is the first key to the solution of the equation of motion influenced by extraterrestrial disturbances. Xu (2004) introduced the so-called disturbance coordinate system by proposing an adjustment model of solar radiation (see Sect. 4.4.2.4 of Xu 2008). However, the coordinate system is not an orthogonal Cartesian one and its axis changes direction with time and therefore the coordinate system is not an inertial one. An approximation of the expression of the solar radiation model is the second key to the solution. The approximation allows the position of the satellite with respect to the Earth to be neglected in the case of solar radiation under special conditions. For a properly selected time interval, the disturbance coordinate system may be considered a frame that has constant rotational relations with respect to the inertial one. In such a case, the coordinate system can be considered approximately "inertial". Then Newton's second law can be used and the orbital disturbance of the solar radiation can be solved. The approximation can be made as precise as required.

The orbits of the satellite can be considered a central motion (Keplerian motion) plus a series of disturbances. According to the order estimation discussed in Sect. 4.4.2.7 of Xu (2008), extraterrestrial perturbations are of second order. These are important for the approximation measure taken during the derivation.

For convenience during later discussions, the definition of the so-called disturbance coordinate system is given again (see Sect. 4.4.2.4 of Xu 2008). The origin is

the geocentre and the three axes are defined by $\vec{r}$ (radial vector of the satellite), $\vec{n}$ (the Sun–satellite unit vector) and $\vec{p}$ (the atmospheric drag unit vector). These three axes are always in the main disturbance directions of the indirect solar radiation (reflected from the Earth's surface), direct solar radiation and atmospheric drag, respectively.

4.4.2 Solutions of Disturbance of Solar Radiation Pressure

Solar radiation pressure is a force caused by sunlight acting on the satellite's surface. The radiation force can be represented as (see (4.70) of Xu 2008)

$$\vec{f}_{\text{solar}} = m\gamma P_{\text{s}} C_{\text{r}} \frac{S}{m} \frac{r_{\text{sun}}^2}{|\vec{r} - \vec{r}_{\text{sun}}|^2} \vec{n}_{\text{sun}}, \tag{26}$$

where

$$\vec{e}_z = -\frac{\vec{r}}{|\vec{r}|}, \qquad \vec{e}_y = \frac{\vec{e}_z \times \vec{n}_{\text{sun}}}{|\vec{e}_z \times \vec{n}_{\text{sun}}|}, \qquad \vec{e}_x = \vec{e}_y \times \vec{e}_z \qquad \text{and} \qquad \vec{n}_{\text{sun}} = \frac{\vec{r} - \vec{r}_{\text{sun}}}{|\vec{r} - \vec{r}_{\text{sun}}|}, \tag{27}$$

where γ is the shadow factor, P_{s} is the luminosity of the Sun, C_{r} is the surface reflectivity, r_{sun} is the geocentric distance of the Sun, S/m is the surface to mass ratio of the satellite and $\vec{r}$ and $\vec{r}_{\text{sun}}$ are the geocentric vectors of the satellite and the Sun. Usually, P_{s} has the value of 4.5605×10^{-6} N/m, C_{r} has values from 1 to 2, 1 is for the complete absorption of the sunlight, and for aluminium, $C_{\text{r}} = 1.95$.

Three Approximations

The solar radiation force vector is pointed from the Sun to the satellite. If the shadow factor is known exactly, and the luminosity of the Sun and the surface reflectivity of the satellite are considered constants, then the length of the solar force vector can be considered a constant because (see (4.73) of Xu 2008)

$$\frac{r_{\text{sun}}^2}{(r_{\text{sun}} + r)^2} \le \frac{r_{\text{sun}}^2}{|\vec{r} - \vec{r}_{\text{sun}}|^2} \le \frac{r_{\text{sun}}^2}{(r_{\text{sun}} - r)^2}. \tag{28}$$

For GPS and GEO satellites there are

$$\frac{r_{\text{sun}}^2}{(r_{\text{sun}} \pm r)^2} = \left(\frac{r_{\text{sun}}}{r_{\text{sun}} \pm r}\right)^2 \approx \left(1 \mp \frac{r}{r_{\text{sun}}} \pm \cdots\right)^2 \approx 1 \mp \frac{2r}{r_{\text{sun}}} \approx 1 \mp 3.5 \times 10^{-4}$$

and

$$\frac{r_{\text{sun}}^2}{(r_{\text{sun}} \pm r)^2} = \left(\frac{r_{\text{sun}}}{r_{\text{sun}} \pm r}\right)^2 \approx \left(1 \mp \frac{r}{r_{\text{sun}}} \pm \cdots\right)^2 \approx 1 \mp \frac{2r}{r_{\text{sun}}} \approx 1 \mp 5.6 \times 10^{-4}, \tag{29}$$

respectively. That is, the solar radiation force vector can be considered approximately a vector, with constant length and changing direction. The approximation

has a precision of better than third order and is precise enough for our purposes. For convenience, this approximation is called the first approximation later on.

The unit solar vector of the satellite $\vec{n}_{\text{sun}}$ can be approximated by

$$\vec{n}_{\text{se}} = \frac{\vec{r}_{\text{se}}}{|\vec{r}_{\text{se}}|}, \tag{30}$$

where index se denotes that the vector is pointing from the Sun to the centre of the Earth. For GPS and GEO satellites the maximal angles between the above two unit vectors are 1.77×10^{-4} and 2.8×10^{-4} rad, respectively. Therefore, such an approximation (called the second approximation) is allowed and is precise enough.

The third approximation is made for suitable time duration of $\Delta t = t'_k - t'_{k-1}$ by

$$\vec{n}_{\text{se}}(t) \approx \vec{n}_{\text{se}}(t_k), \quad t_k = (t'_k + t'_{k-1})/2, \quad t \in [t'_{k-1}, t'_k]. \tag{31}$$

The discrete vector in this equation may be called an average vector of the time duration Δt. For $\Delta t = 5$ min, the third approximation has a precision of 3×10^{-5} rad.

Note that the order of the solar radiation disturbance on a GPS satellite is about 50 m. For GPS satellite, all the three approximations will lead to a precision of millimetre level. For the other satellite, the precision of the approximations should be individually estimated.

Discretisation and Solution

Denote the satellite period as T and shadow access and exit points as t_{a} and t_{e}, respectively. The local noon is selected as the starting point of counting (see Fig. 4.4). A so-called sign function can be defined as

$$\delta(t) = \begin{cases} 1 & 0 \le t < T/2 \\ -1 & T/2 \le t \le T \end{cases}. \tag{32}$$

The sign function shows that the solar radiation accelerates the satellite during the first half period and decelerates it during the second half period with respect to

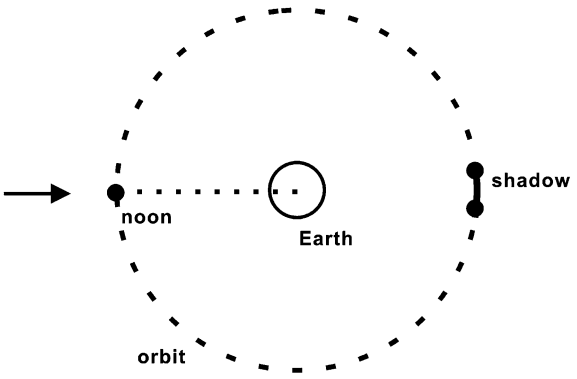


Fig. 4.4 Solar radiation pressure

the nominal motion of the satellite. Then the duration of one period of 0–T can be equally divided by Δt, i.e. by $t'_0, t'_1, \ldots, t'_k, \ldots, T$. The acceleration of the solar radiation of (26) is then discretised as

$$\vec{a}_{\text{solar}}(t) = \gamma P_{\text{s}} C_{\text{r}} \frac{S}{m} \vec{n}_{\text{se}}(t_k). \tag{33}$$

The disturbed velocity caused by the solar radiation is then

$$\vec{v}_{\text{solar}}(t) = \sum_{i=1}^{k} \gamma P_{\text{s}} C_{\text{r}} \frac{S}{m} \vec{n}_{\text{se}}(t_i) \delta(t_i) \Delta t. \tag{34}$$

It is obvious that the disturbed velocity of the satellite is not zero during the passing of the shadow. The disturbed position caused by the solar radiation is then

$$\vec{\rho}_{\text{solar}}(t) = \sum_{j=1}^{k} \vec{v}_{\text{solar}}(t_j) \Delta t. \tag{35}$$

Equation (35) is the solution of the solar radiation disturbance on the orbit of the satellite.

Properties of the Solution

The integration (or summation) of the acceleration of the solar radiation within a period T is nearly zero. However, the position disturbed by the solar radiation during a period T is not zero. In other words, the disturbance of the solar radiation has non-conservative behaviour. The disturbance may not be a periodic function of the orbit. The parameters of the force model, if they are not well known, can be determined using the expressions of the solution.

4.4.2.1 Solutions via Gaussian Perturbed Equations

Gaussian Perturbed Equations

Equation (33) is the approximated solar radiation force (acceleration) vector with constant length, which can be written as

$$\vec{f}_{\text{solar}}(t) = m \gamma P_{\text{s}} C_{\text{r}} \frac{S}{m} \vec{n}_{\text{se}}(t) \tag{36}$$

or

$$\vec{f}_{\text{solar}}(t) = \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \xi \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}, \tag{37}$$

where solar-Earth unit vector (30) in ECSF frame can be computed by the theory given in Sect. 4.7.8 of Xu (2008); ξ represents the constant length of the solar radiation force vector.

The force vector in the ECSF frame can be transformed to the orbital coordinate system (see (13)) using

$$\begin{pmatrix} f_{\text{r}} \\ f_{\alpha} \\ f_{\text{h}} \end{pmatrix} = R_3(f)R_3(\omega)R_1(i)R_3(\Omega) \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}, \tag{38}$$

where

$$R_3(\omega)R_1(i)R_3(\Omega) = \begin{pmatrix} \cos\omega\cos\Omega - \sin\omega\cos i\sin\Omega & \cos\omega\sin\Omega + \sin\omega\cos i\cos\Omega & \sin\omega\sin i \\ -\sin\omega\cos\Omega - \cos\omega\cos i\sin\Omega & -\sin\omega\sin\Omega + \cos\omega\cos i\cos\Omega & \cos\omega\sin i \\ \sin i\sin\Omega & -\sin i\cos\Omega & \cos i \end{pmatrix}$$

Denote these elements of the matrix with R_{ij} and

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \xi \begin{pmatrix} R_{11}n_x + R_{12}n_y + R_{13}n_z \\ R_{21}n_x + R_{22}n_y + R_{23}n_z \\ R_{31}n_x + R_{32}n_y + R_{33}n_z \end{pmatrix}, \tag{39}$$

then one has

$$\begin{pmatrix} f_{\text{r}} \\ f_{\alpha} \\ f_{\text{h}} \end{pmatrix} = R_3(f) \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} n_1\ \cos f + n_2\ \sin f \\ -n_1\ \sin f + n_2\ \cos f \\ n_3 \end{pmatrix}. \tag{40}$$

There are relations (see (4.23) of Xu 2008)

$$\frac{(1-e^2)}{1+e\cos f} = (1 - e\cos E) \tag{41}$$

and

$$\cos E = \frac{(e+\cos f)}{1+e\cos f}. \tag{42}$$

Putting all these formulas into (17), the Gaussian disturbed equations are

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{2}{n\sqrt{1-e^2}}\left[\begin{array}{l} e\ \sin f(n_1\ \cos f + n_2\ \sin f) \\ +(1+e\ \cos f)(-n_1\ \sin f + n_2\ \cos f)\end{array}\right] \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{na}\left[\begin{array}{l} \sin f(n_1\ \cos f + n_2\ \sin f) \\ +\left(\dfrac{e+\cos f}{1+e\ \cos f}+\cos f\right)(-n_1\ \sin f + n_2\ \cos f)\end{array}\right] \\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}}{nae}\left[\begin{array}{l} -\cos f(n_1\ \cos f + n_2\ \sin f) \\ +\dfrac{2+e\ \cos f}{1+e\ \cos f}\sin f(-n_1\ \sin f + n_2\ \cos f)\end{array}\right] - \cos i\frac{\mathrm{d}\Omega}{\mathrm{d}t} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}\ \cos u}{na(1+e\ \cos f)}n_3 \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \frac{\sqrt{1-e^2}\sin u}{na\ \sin i(1+e\ \cos f)}n_3 \\
\frac{\mathrm{d}M}{\mathrm{d}t} &= -\frac{1-e^2}{nae}\left[\begin{array}{l} -\left(\cos f - \dfrac{2e}{1+e\ \cos f}\right)(n_1\ \cos f + n_2\ \sin f) \\ +\dfrac{2+e\ \cos f}{1+e\ \cos f}\sin f(-n_1\ \sin f + n_2\ \cos f)\end{array}\right]
\end{aligned} \quad . \tag{43}$$

Characters of Gaussian Perturbed Equations

1. There exist long and short periodic perturbations.

Note that

$$\begin{aligned}
\sin^2 f &= \frac{1-\cos^2 f}{2} \\
\cos^2 f &= \frac{1+\cos^2 f}{2}
\end{aligned} \tag{44}$$

$$\frac{1}{1+e\ \cos f} \approx 1 - e\ \cos f + \cdots \tag{45}$$

and

$$\begin{aligned}
\cos u &= \cos\omega\ \cos f - \sin\omega\ \sin f \\
\sin u &= \cos\omega\ \sin f + \cos f\ \sin\omega
\end{aligned} \quad . \tag{46}$$

Obviously, all six Gaussian perturbed equations include the long periodic term perturbations, which are formed by terms without f (in other words, constant terms are created by terms of $\sin^2 f$ and $\cos^2 f$), and the remaining terms are short periodic terms. Remember that by integration variable transformation from t to f or M for solving the short periodic $\overline{C}_{20}$ perturbations, long periodic terms will also be created (see Sect. 4.5.2 of Xu 2008). Therefore, no effort will be made to separate the long and short periodic terms.

2. Concerning time variable (n_x, n_y and n_z) in (43) the variables n_1, n_2 and n_3 are functions of (ω, Ω, i) and (n_x, n_y, n_z). (ω, Ω, i) are long periodic variables and they are considered constants in short periodic integrations. However, the unit vector (n_x, n_y, n_z) of solar-Earth is also time variable. In the discussion in Sect. 4.4.3 the (n_x, n_y, n_z) can be considered constants within 5 min. The maximum change of the unit vector around its average is ca. 0.5°/day, that is, the maximum of change rate is about 0.0086 rad/day. In other words, the unit vector (n_x, n_y, n_z) can be represented by an average plus a drift term, and the drift term compared with the average term is about one order smaller and in some cases is allowed to be neglected. As soon as the vector (n_x, n_y, n_z) is considered constant, (43) can be solved by integration as shown in Chaps. 5 and 6 of Xu (2008).

In cases where change of the unit vector is not allowed to be neglected, the integration interval has to be made shorter so that the assumption will be valid and then the integrated solution should be summated to obtain the complete solutions.

Solutions of Gaussian Perturbed Equations

For simplifying the disturbed equations, denote

$$n_4 = \frac{n_3 e}{\sin i}, \quad n_5 = n_4 \cos i, \quad n_6 = \frac{n_5}{n_4}, \quad g_1 = \frac{2}{n\sqrt{1-e^2}}, \quad g_2 = \frac{\sqrt{1-e^2}}{na},$$
$$g_3 = g_2, \quad g_4 = \frac{\sqrt{1-e^2}}{nae}, \quad g_5 = g_4, \quad g_6 = -\frac{1-e^2}{nae}. \tag{47}$$

Omitting the factors g_j ($j = 1, \ldots, 6$) in the disturbing equations (of course, after the equations are solved, the factors shall be multiplied back), one has

$$\begin{aligned}
\frac{da}{dt} &= \begin{bmatrix} e \sin f(n_1 \cos f + n_2 \sin f) \\ +(1 + e \cos f)(-n_1 \sin f + n_2 \cos f) \end{bmatrix} \\
\frac{de}{dt} &= \begin{bmatrix} \sin f(n_1 \cos f + n_2 \sin f) \\ +\left(\dfrac{e + \cos f}{1 + e \cos f} + \cos f\right)(-n_1 \sin f + n_2 \cos f) \end{bmatrix} \\
\frac{d\omega}{dt} &= \begin{bmatrix} -\cos f(n_1 \cos f + n_2 \sin f) \\ +\dfrac{2 + e \cos f}{1 + e \cos f} \sin f(-n_1 \sin f + n_2 \cos f) \end{bmatrix} - \frac{d\Omega}{dt} n_6 \\
\frac{di}{dt} &= \frac{\cos u}{1 + e \cos f} n_3 \\
\frac{d\Omega}{dt} &= \frac{\sin u}{1 + e \cos f} n_4 \\
\frac{dM}{dt} &= \begin{bmatrix} -\left(\cos f - \dfrac{2e}{1 + e \cos f}\right)(n_1 \cos f + n_2 \sin f) \\ +\dfrac{2 + e \cos f}{1 + e \cos f} \sin f(-n_1 \sin f + n_2 \cos f) \end{bmatrix}
\end{aligned} \tag{48}$$

Let

$$\omega_1 = \omega + n_6\Omega, \quad M_1 = -M + \omega_1, \quad \frac{1}{1 + e\,\cos f} \approx 1 - e\,\cos f, \tag{49}$$

and (48) can be further simplified as

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \begin{bmatrix} e\,\sin f(n_1\,\cos f + n_2\,\sin f) \\ +(1 + e\,\cos f)(-n_1\,\sin f + n_2\,\cos f) \end{bmatrix} \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \begin{bmatrix} \sin f(n_1\,\cos f + n_2\,\sin f) \\ +(e + (2 - e^2)\cos f - e\,\cos^2 f)(-n_1\,\sin f + n_2\,\cos f) \end{bmatrix} \\
\frac{\mathrm{d}\omega_1}{\mathrm{d}t} &= \begin{bmatrix} -\cos f(n_1\,\cos f + n_2\,\sin f) \\ +(2 + e\,\cos f)(1 - e\,\cos f)\sin f(-n_1\,\sin f + n_2\,\cos f) \end{bmatrix} \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \cos u(1 - e\,\cos f)n_3 \\
\frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \sin u(1 - e\,\cos f)n_4 \\
\frac{\mathrm{d}M_1}{\mathrm{d}t} &= \left[-2\,e(1 - e\,\cos f)(n_1\,\cos f + n_2\,\sin f)\right].
\end{aligned} \tag{50}$$

Simplified Gaussian perturbed equations (50) may be solved using symbolic computational software. The infinite integrations of the differential equations can be represented by

$$(\Delta\sigma_j(M))_M = b_j \left(d_j(\omega, \Omega)M + \sum_{k=1}^{16} A_{jk}\,\cos\,kM + \sum_{k=1}^{16} B_{jk}\,\sin\,kM \right), \tag{51}$$

where j is the index of Keplerian elements, b_j includes the omitted factors g_j and the factor caused by the variable transformation from t to M (see (5.24) of Xu 2008) as well as the factors h_j given below:

$$\begin{aligned}
&h_1 = (1,152 \times 210)^{-1}, \quad h_2 = (55,296 \times 2,310)^{-1}, \quad h_3 = h_1 \\
&h_4 = h_1, \quad h_5 = (2654,208 \times 60,060)^{-1}, \quad h_6 = (576 \times 210)^{-1},
\end{aligned} \tag{52}$$

where h_j factors are introduced to simplify the derivations of (51). The first term on the right-hand side of (51) is symbolic and represents the long periodic perturbation of

$$\int d_j(\omega, \Omega)\,\mathrm{d}M. \tag{53}$$

dM can be transformed to d$(n\omega+m\Omega)$ depending on the form of d_j according to (53).

Formulas of d_j, A_{jk} and B_{jk} are given in detail in Xu (2008).

Properties of the Solution

Disturbances of the solar radiation consist of both the long periodic and short periodic terms. The orientation of the orbital ellipse is subjected to higher frequency disturbance than that of the other Keplerian elements.

4.4.3 Solutions of Disturbance of Atmospheric Drag

Atmospheric drag, caused by the air, is the disturbance force acting on the satellite's surface. Air drag force can be represented as (see (4.75) of Xu 2008)

$$\vec{f}_{\text{drag}} = -m\frac{1}{2}\left(\frac{C_{\text{d}}S}{m}\right)\sigma\left|\dot{\vec{r}} - \dot{\vec{r}}_{\text{air}}\right|^2 \vec{n}_{\text{a}}, \quad \vec{n}_{\text{a}} = \frac{\dot{\vec{r}} - \dot{\vec{r}}_{\text{air}}}{\left|\dot{\vec{r}} - \dot{\vec{r}}_{\text{air}}\right|}, \tag{54}$$

where S is the cross section (or effective area) of the satellite, C_{d} is the drag factor, m is the mass of the satellite, $\dot{\vec{r}}$ and $\dot{\vec{r}}_{\text{air}}$ are the geocentric velocity vectors of the satellite and the atmosphere, respectively, and σ is the density of the atmosphere. Usually, S has a value of 1/4 of the outer surface area of the satellite, and C_{d} has labour values of 2.2 ± 0.2. The unit vector $\vec{n}_a$ is the direction of the air drag force. For CHAMP satellite, with an orbit height of 400 km, the air drag force unit vector $\vec{n}_{\text{a}}$ changes its direction about 1.2×10^{-3} rad/s. The amount of $\left|\dot{\vec{r}} - \dot{\vec{r}}_{\text{air}}\right|^2$ changes slower than the direction. In such a case the acceleration of the air drag can be discretised by

$$\vec{a}_{\text{drag}} = -\frac{1}{2}\left(\frac{C_{\text{d}}S}{m}\right)\sigma\left|\dot{\vec{r}}(t_k) - \dot{\vec{r}}_{\text{air}}(t_k)\right|^2 \vec{n}_{\text{a}}(t_k). \tag{55}$$

The disturbed velocity caused by the atmospheric drag is then

$$\vec{v}_{\text{air drag}}(t) = \sum_{i=1}^{k} -\frac{C_{\text{d}}S}{2m}\sigma\left|\dot{\vec{r}}(t_i) - \dot{\vec{r}}_{\text{air}}(t_i)\right|^2 \vec{n}_{\text{a}}(t_i)\Delta t. \tag{56}$$

The disturbed position caused by the solar radiation is then

$$\vec{\rho}_{\text{air}}(t) = \sum_{j=1}^{k} \vec{v}_{\text{air drag}}(t_j)\Delta t. \tag{57}$$

Equation (57) is the solution of the solar radiation disturbance on the orbit of the satellite.

For all satellites, with an orbit height higher than 1,000 km, the atmospheric drag is nearly zero; therefore this effect does not need to be taken into account.

4.4.3.1 Solutions via Gaussian Perturbed Equations

Air Drag Force Vector for Gaussian Perturbed Equations

Air drag force is given in (54) (using ξ to represent the coefficient part of the air drag force vector)

$$\vec{f}_{\text{drag}} = \xi \left| \dot{\vec{r}} - \dot{\vec{r}}_{\text{air}} \right| (\dot{\vec{r}} - \dot{\vec{r}}_{\text{air}}). \tag{58}$$

Using (38) the air drag force vector can be rotated from the ECSF to the orbital coordinate frame by

$$\begin{pmatrix} f_{\text{r}} \\ f_{\alpha} \\ f_{\text{h}} \end{pmatrix} = R_3(f)R_3(\omega)R_1(i)R_3(\Omega) \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}. \tag{59}$$

Satellite position and velocity vectors in orbital frame are (see (3.41) and (3.42) of Xu 2008)

$$\vec{q} = \begin{pmatrix} a(\cos E - e) \\ a\sqrt{1-e^2}\,\sin E \\ 0 \end{pmatrix} = \begin{pmatrix} r\cos f \\ r\sin f \\ 0 \end{pmatrix}, \tag{60}$$

$$\dot{\vec{q}} = \begin{pmatrix} -\sin E \\ \sqrt{1-e^2}\cos E \\ 0 \end{pmatrix} \frac{na}{1-e\cos E} = \begin{pmatrix} -\sin f \\ e+\cos f \\ 0 \end{pmatrix} \frac{na}{\sqrt{1-e^2}}. \tag{61}$$

They can be rotated from the orbital frame to the ECSF frame (see (3.43) of Xu 2008):

$$\begin{pmatrix} \vec{r} \\ \dot{\vec{r}} \end{pmatrix} = R_3(-\Omega)R_1(-i)R_3(-\omega) \begin{pmatrix} \vec{q} \\ \dot{\vec{q}} \end{pmatrix}. \tag{62}$$

Air velocity in the ECSF frame is given in (4.76) of Xu (2008):

$$\dot{\vec{r}}_{\text{air}} = k\vec{\omega}_{\text{e}} \times \vec{r} = k\omega_{\text{e}} \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix} = k\omega_{\text{e}} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = k\omega_{\text{e}} R_4 \vec{r}, \tag{63}$$

where ω_{e} is the angle velocity of the Earth's rotation. Thus in the ECSF frame there is

$$\ddot{\vec{r}} - \ddot{\vec{r}}_{\text{air}} = R_3(-\Omega)R_1(-i)R_3(-\omega)\ddot{\vec{q}} - k\omega_e R_4 R_3(-\Omega)R_1(-i)R_3(-\omega)\vec{q}. \tag{64}$$

Denote the following matrix as R:

$$R_3(\omega)R_1(i)R_3(\Omega)R_4R_3(-\Omega)R_1(-i)R_3(-\omega) = R_3(\omega)R_1(i)R_4R_1(-i)R_3(-\omega)$$
$$= R_3(\omega)\begin{pmatrix} 0 & -\cos i & \sin i \\ \cos i & 0 & 0 \\ -\sin i & 0 & 0 \end{pmatrix}R_3(-\omega) = \begin{pmatrix} 0 & -\cos i & \sin i\cos\omega \\ \cos i & 0 & -\sin i\sin\omega \\ -\sin i\cos\omega & \sin i\sin\omega & 0 \end{pmatrix} \tag{65}$$

and note that the length of a vector is invariable under rotational transformations; one has

$$\begin{pmatrix} f_\mathrm{r} \\ f_\alpha \\ f_\mathrm{h} \end{pmatrix} = \xi\left|\dot{\vec{r}} - \dot{\vec{r}}_\mathrm{air}\right| R_3(f)(\dot{\vec{q}} - k\omega_\mathrm{e}R\vec{q}) = \xi\left|\dot{\vec{q}} - k\omega_\mathrm{e}R\vec{q}\right| R_3(f)(\dot{\vec{q}} - k\omega_\mathrm{e}R\vec{q}). \tag{66}$$

The force vector (66) is represented completely in Keplerian elements.

Gaussian Perturbed Equations and the Solutions

The air drag force vector (66) has to be further simplified. Denote the elements of the matrix R with R_{ij}; then one has approximately

$$\begin{aligned}\dot{\vec{q}} - k\omega_\mathrm{e}R\vec{q} &= \frac{na}{\sqrt{1-e^2}}\begin{pmatrix} -\sin f \\ e+\cos f \\ 0 \end{pmatrix} - k\omega R(1-e^2)\begin{pmatrix} \cos f \\ \sin f \\ 0 \end{pmatrix}(1 - e\cos f) \\ &= \begin{pmatrix} b_{11}\sin f + b_{13}\sin f\cos f \\ b_{22}\cos f + b_{24}\cos^2 f + b_{25} \\ b_{31}\sin f + b_{32}\cos f + b_{33}\sin f\cos f + b_{34}\cos^2 f \end{pmatrix},\end{aligned} \tag{67}$$

where coefficients b_{ij} can be obtained by comparison.

For convenience, the simplified Gaussian disturbed equations of motion can be written as shown below (see (50), (49) and (47)):

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= \left[e\sin f\cdot f_\mathrm{r} + (1+e\cos f)\cdot f_\alpha\right] \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= \left[\sin f\cdot f_\mathrm{r} + (e\sin^2 f + (2-e^2)\cos f)\cdot f_\alpha\right] \\ \frac{\mathrm{d}\omega_1}{\mathrm{d}t} &= \left[-\cos f\cdot f_\mathrm{r} + (2+e\cos f)(1-e\cos f)\sin f\cdot f_\alpha\right] \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \cos u(1-e\cos f)\cdot f_\mathrm{h} \\ \frac{\mathrm{d}\Omega}{\mathrm{d}t} &= \sin u(1-e\cos f)\frac{e}{\sin i}\cdot f_\mathrm{h} \\ \frac{\mathrm{d}M_1}{\mathrm{d}t} &= \left[-2e(1-e\cos f)\cdot f_\mathrm{r}\right] \end{aligned} \tag{68}$$

Putting the air drag force vector and other mathematical relations into the simplified Gaussian disturbed equations (68), the equations could be solved (see Xu et al. 2010b).

4.4.4 Solutions of Disturbance of the Sun

The solutions of the disturbance of the Sun may be similarly derived by the discretisation demonstrated in Sect. 4.4.2. However, analytic solutions are preferred in theoretical and practical aspects.

Potential Function of the Sun

The disturbance forces of multiple point masses acting on the satellite are (see (4.50) of Xu 2008)

$$\vec{f}_{\text{mul}} = -m \sum_j Gm(j) \left[\frac{\vec{r} - \vec{r}_{m(j)}}{\left|\vec{r} - \vec{r}_{m(j)}\right|^3} + \frac{\vec{r}_{m(j)}}{r_{m(j)}^3} \right], \tag{69}$$

where $Gm(j)$ are the gravitational constants of the Sun and the Moon as well as the planets. The disturbance acceleration of the Sun is then

$$\begin{aligned} \vec{f}_{\text{s}} &= -m\mu_{\text{s}} \left(\frac{1}{|\vec{r} - \vec{r}_{\text{s}}|^2} \vec{n}_{\text{ss}} + \frac{1}{|\vec{r}_{\text{s}}|^2} \vec{n}_{\text{s}} \right), \quad \vec{n}_{\text{ss}} = \frac{\vec{r} - \vec{r}_{\text{s}}}{|\vec{r} - \vec{r}_{\text{s}}|}, \quad \vec{n}_{\text{s}} = \frac{\vec{r}_{\text{s}}}{|\vec{r}_{\text{s}}|} \\ &= -m\mu_{\text{s}} \left[\frac{r}{|\vec{r} - \vec{r}_{\text{s}}|^3} \vec{n}_{\text{r}} + \left(\frac{1}{r_{\text{s}}^2} - \frac{r_{\text{s}}}{|\vec{r} - \vec{r}_{\text{s}}|^3} \right) \vec{n}_{\text{s}} \right], \quad \vec{n}_{\text{r}} = \frac{\vec{r}}{r} \end{aligned} . \tag{70}$$

The unit vectors $\vec{n}_{\text{ss}}$, $\vec{n}_{\text{s}}$ and $\vec{n}_{\text{r}}$ represent the vector from the Sun to the satellite, the geocentric vectors of the Sun and satellite, respectively. The force vector of the Sun is a vector summated from two vectors in directions of $\vec{n}_{\text{r}}$ and $\vec{n}_{\text{s}}$ (see Fig. 4.5). According to the geometric relations of the vectors one has (cf. Fig. 4.6)

$$\begin{aligned} &\vec{r} = \vec{r}_{\text{s}} + (\vec{r} - \vec{r}_{\text{s}}) \\ &|\vec{r} - \vec{r}_{\text{s}}|^2 = r_{\text{s}}^2 + r^2 - 2rr_{\text{s}} \cos\alpha\,, \\ &\cos\alpha = \vec{n}_{\text{r}} \cdot \vec{n}_{\text{s}} \end{aligned} \tag{71}$$

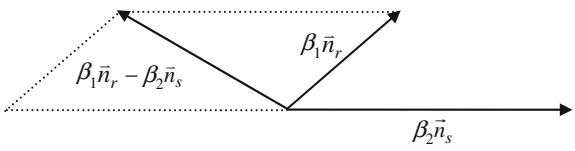


Fig. 4.5 Disturbance force vector and its two components (two β with indices are lengths)

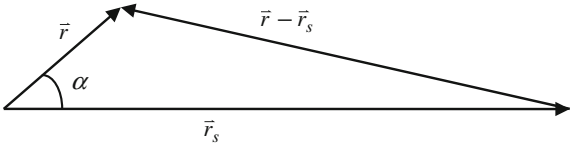


Fig. 4.6 Triangle relationship between the geocentric vectors of the satellite and the Sun

where α is the angle between the geocentric vectors of the satellite and the Sun and $\cos\alpha$ is the inner product of the two unit vectors. Using the mathematic formula

$$\sqrt[q]{(1+x)^p} = 1 + \frac{p}{q}x + \frac{p(p-q)}{q\cdot 2q}x^2 + \cdots \quad |x| \le 1,\ p > 0 \quad \text{or} \quad q > 0, \tag{72}$$

the lengths of the two vectors in (70) can be approximated by

$$\begin{aligned} \frac{r}{|\vec{r}-\vec{r}_s|^3} &\approx \frac{r}{r_s^3}\left(1 + 3\frac{r}{r_s}\cos\alpha\right) \\ \frac{1}{r_s^2} - \frac{r_s}{|\vec{r}-\vec{r}_s|^3} &\approx \frac{r}{r_s^3}\left(-3\cos\alpha - 1.5\frac{r}{r_s}(5\cos^2\alpha - 1)\right) \end{aligned}. \tag{73}$$

The geocentric distance rates of GPS and GEO satellites to the Sun are about 1.77×10^{-4} and 2.8×10^{-4}, respectively. Therefore, for all satellites which are lower than the GEO satellite, above two approximations in (73) are precise enough to take only the first term into account and then one has

$$\vec{f}_s = -m\mu_s\left(\frac{r}{r_s^3}\vec{n}_r - \frac{3r\cos\alpha}{r_s^3}\vec{n}_s\right). \tag{74}$$

The potential function of the disturbing force of the Sun (74) is then

$$V_s = -m\mu_s\frac{r}{r_s^3} + m\mu_s\frac{3}{2r_s^3}(\vec{n}_s\cdot\vec{r})^2. \tag{75}$$

The correctness of the potential function can be verified directly by making gradient operation on (75) and comparing the results with the force vector (74). The computation shows that for GPS satellite within 3 days the force vector (70) and its approximation (74) have differences of 3×10^{-10} (that is one order smaller than that of the computed by using (70)).

Disturbed Equation of Motion and the Solutions

Denote the first term of the potential function (75) as V_{s1} and note that V_{s1} is the function of the three Keplerian elements (a,M,e). The derivatives of the potential function with respect to Keplerian elements are then

$$\frac{\partial V_{s1}}{\partial a} = \frac{\partial V_{s1}}{\partial r}\frac{\partial r}{\partial a} = \frac{1}{a}V_{s1}, \quad \frac{\partial V_{s1}}{\partial \Omega} = \frac{\partial V_{s1}}{\partial i} = \frac{\partial V_{s1}}{\partial \omega} = 0,$$
$$\frac{\partial V_{s1}}{\partial e} = \frac{V_{s1}}{r}\frac{\partial r}{\partial e} = \frac{-a\cos f}{r}V_{s1}$$

and

$$\frac{\partial V_{s1}}{\partial M} = \frac{V_{s1}}{r}\frac{\partial r}{\partial M} = \frac{ae\,\sin f}{r\sqrt{1-e^2}}V_{s1}. \tag{76}$$

Substituting the above derivatives and V_{s1} into the equation of motion (11), one has

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}t} &= \frac{-2m\mu_s}{nr_s^3}\frac{e\,\sin f}{\sqrt{1-e^2}},\\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \frac{-m\mu_s\sqrt{1-e^2}}{na}\frac{\sin f}{r_s^3},\\
\frac{\mathrm{d}\omega}{\mathrm{d}t} &= \frac{m\mu_s\sqrt{1-e^2}}{nae}\frac{\cos f}{r_s^3},\\
\frac{\mathrm{d}i}{\mathrm{d}t} &= 0, \quad \frac{\mathrm{d}\Omega}{\mathrm{d}t} = 0,
\end{aligned} \tag{77}$$

$$\begin{aligned}
\frac{\mathrm{d}M}{\mathrm{d}t} &= \frac{2}{na}\frac{m\mu_s}{r_s^3}\frac{(1-e^2)}{1+e\,\cos f} - \frac{1-e^2}{nae}\frac{m\mu_s\,\cos f}{r_s^3}\\
&= \frac{m\mu_s(1-e^2)}{naer_s^3}\left(2e\frac{1}{1+e\,\cos f} - \cos f\right).\\
&= \frac{m\mu_s(1-e^2)}{naer_s^3}(2e-(1+2e^2)\cos f)
\end{aligned}$$

Geocentric distance of the Sun can be considered a daily constant. There are (see Liu and Zhao 1979 and (5.22) of Xu 2008)

$$\begin{aligned}
\sin f &= \left(1-\frac{7}{8}e^2\right)\sin M + e\left(1-\frac{7}{6}e^2\right)\sin 2M + \frac{9}{8}e^2\sin 3M + \frac{4}{3}e^3\sin 4M,\\
\cos f + e &= \left(1-\frac{9}{8}e^2\right)\cos M + e\left(1-\frac{4}{3}e^2\right)\cos 2M + \frac{9}{8}e^2\cos 3M + \frac{4}{3}e^3\cos 4M.
\end{aligned} \tag{78}$$

Denote

$$\begin{aligned}
\delta S &= \int \sin f\,\mathrm{d}t = \int \sin f\left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_0^{-1}\mathrm{d}M = \left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_0^{-1}\cdot\\
&\left(-\left(1-\frac{7}{8}e^2\right)\cos M - \frac{e}{2}\left(1-\frac{7}{6}e^2\right)\cos 2M - \frac{3}{8}e^2\cos 3M - \frac{1}{3}e^3\cos 4M\right)\\
\delta C &= \int(\cos f + e)\mathrm{d}t = \int(\cos f + e)\left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_0^{-1}\mathrm{d}M = \left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_0^{-1}\cdot\\
&\left(\left(1-\frac{9}{8}e^2\right)\sin M + \frac{e}{2}\left(1-\frac{4}{3}e^2\right)\sin 2M + \frac{3}{8}e^2\sin 3M + \frac{1}{3}e^3\sin 4M\right).
\end{aligned} \tag{79}$$

Then the solutions are

$$\Delta a = \frac{-2m\mu_{\rm s}}{nr_{\rm s}^3}\frac{e}{\sqrt{1-e^2}}\delta S,$$

$$\Delta e = \frac{-m\mu_{\rm s}\sqrt{1-e^2}}{nar_{\rm s}^3}\delta S,$$

$$\Delta\omega = \frac{m\mu_{\rm s}\sqrt{1-e^2}}{naer_{\rm s}^3}(-et+\delta C),$$

$$\Delta i = \text{const.}, \quad \Delta\Omega = \text{const.}, \tag{80}$$

$$\Delta M = \frac{m\mu_{\rm s}(1-e^2)}{naer_{\rm s}^3}(e(3+2e^2)t-(1+2e^2)\delta C).$$

The orbital parameters ω and M are partly linearly perturbed by the Sun.

Denote the second term of the potential function (75) as $V_{\rm s2}$ and note (cf. (62))

$$\begin{aligned}
\vec{r} &= R_3(-\Omega)R_1(-i)R_3(-\omega)r\begin{pmatrix}\cos f\\ \sin f\\ 0\end{pmatrix} = rR\begin{pmatrix}\cos f\\ \sin f\\ 0\end{pmatrix}\\
&= r\begin{pmatrix}\cos\Omega\cos i\cos\omega-\sin\Omega\sin\omega & -\cos\Omega\cos i\sin\omega-\sin\Omega\sin\omega & \cos\Omega\sin i\\ \sin\Omega\sin i\cos\omega+\cos\Omega\sin\omega & -\sin\Omega\sin i\sin\omega+\cos\Omega\cos\omega & \sin\Omega\sin i\\ -\sin i\cos\omega & \sin i\sin\omega & \cos i\end{pmatrix}\begin{pmatrix}\cos f\\ \sin f\\ 0\end{pmatrix}\\
&= r\begin{pmatrix}R_{11}\cos f+R_{12}\sin f\\ R_{21}\cos f+R_{22}\sin f\\ R_{31}\cos f+R_{32}\sin f\end{pmatrix}
\end{aligned} \tag{81}$$

$$\vec{n}_{\rm s} = \begin{pmatrix}n_{\rm s1}\\ n_{\rm s2}\\ n_{\rm s3}\end{pmatrix},$$

$$r\cos\alpha = \vec{n}_{\rm s}\cdot\vec{r} = r(A\cos f + B\sin f), \tag{82}$$

where coefficients are

$$\begin{aligned}
A &= n_{\rm s1}R_{11}+n_{\rm s2}R_{21}+n_{\rm s3}R_{31}\\
B &= n_{\rm s1}R_{12}+n_{\rm s2}R_{22}+n_{\rm s3}R_{32},
\end{aligned} \tag{83}$$

where R with a single index k is rotational matrix around the axis k; R with two indices are elements of the total rotational matrix R and they are triangle functions of (Ω, i, ω); solar-related elements are denoted with index s; n_{sk} is the k-th component of the geocentric unit vector of the Sun; A and B are functions of sinus and cosines of (Ω, i, ω) as well as components of the geocentric solar unit vector. Then one has derivatives of the second-term potential function with respect to Keplerian elements

$$\frac{\partial V_{s2}}{\partial \sigma_j} = m\mu_s \frac{3}{r_s^3} r \cos\alpha \frac{\partial r \cos\alpha}{\partial \sigma_j}, \tag{84}$$

where σ with index j denotes the j-th Keplerian element. Explicitly the partial derivatives are (cf. (4.24) in Xu 2008)

$$\frac{\partial V_{s2}}{\partial a} = m\mu_s \frac{3}{r_s^3} \frac{r^2}{a} \cos^2\alpha, \tag{85}$$

$$\frac{\partial V_{s2}}{\partial \Omega} = m\mu_s \frac{3}{r_s^3} r^2 \cos\alpha (A_\Omega \cos f + B_\Omega \sin f), \tag{86}$$

$$\frac{\partial V_{s2}}{\partial i} = m\mu_s \frac{3}{r_s^3} r^2 \cos\alpha (A_i \cos f + B_i \sin f), \tag{87}$$

$$\frac{\partial V_{s2}}{\partial \omega} = m\mu_s \frac{3}{r_s^3} r^2 \cos\alpha (A_\omega \cos f + B_\omega \sin f), \tag{88}$$

$$\frac{\partial V_{s2}}{\partial e} = m\mu_s \frac{3}{r_s^3} r \cos\alpha \left(\cos\alpha(-a\cos f) + r(-A\sin f + B\cos f)\frac{2+e\cos f}{1-e^2}\sin f\right), \tag{89}$$

and

$$\frac{\partial V_{s2}}{\partial M} = m\mu_s \frac{3}{r_s^3} \cos\alpha \left(\frac{ae}{\sqrt{1-e^2}} r \cos\alpha \sin f + (-A\sin f + B\cos f) a^2\sqrt{1-e^2}\right) \tag{90}$$

where A and B with indices of (Ω, i, ω) are partial derivatives of A and B with respect to the related indices. Putting (85), (86), (87), (88), (89) and (90) into (11) and taking the following approximation (terms with order e^2 are neglected) into account (cf. (72))

$$r = \frac{a(1-e^2)}{1+e\cos f} \approx a(1 - e\cos f + \cdots), \tag{91}$$

$$r^2 = \frac{a^2(1-e^2)^2}{(1+e\cos f)^2} \approx a^2(1 - 2e\cos f + \cdots),$$

$$\frac{1}{\sin i} = \frac{1}{\sqrt{1-\cos^2 i}} \approx 1 + \frac{1}{2}\cos^2 i + \frac{3}{8}\cos^4 i + \frac{5}{16}\cos^6 i + \cdots$$

the Lagrangian equations of satellite motion can be transformed in terms of $\cos^m f \sin^k f$ and then in terms of multiplications of $\{\cos mM,\ \cos kM,\ \sin mM,\ \sin kM\}$; and at the end they can be reduced to functional series of $\{\cos mM,\ \sin kM\}$; which can be integrated with respect to M to obtain the short periodic terms solutions. The terms nothing to do with M are long periodic terms which can be reduced to triangle functions of $(k\Omega + mi + l\omega)$ and long (non-periodic) terms. Where k, m and l, that are integers. The transformation can be carried out by using mathematic symbol operation software such as Mathematica or Maple and the solutions have the form of

$$\Delta\sigma_j = d_j + \sum_{k=1,\ldots} b_{jk}\cos\, kM + c_{jk}\sin\, kM,$$

$$d_j = L_j + \sum_{k,m,l} g_{jkml}\cos(k\Omega + mi + l\omega) + h_{jkml}\sin(k\Omega + mi + l\omega), \qquad (92)$$

where *d, b, c* are functions of (Ω, *i, ω, a, e*) and *g, h, L* are functions of (*a, e*). All terms of *b* and *c* are short periodic terms; all terms of *g* and *h* are long periodic terms; and *L* terms are the long-term effects. Due to the length of the formulas, *b* and *c*, *g* and *h* will not be given here (for detail see Xu et al. 2010a). Selecting the integration interval as $k2\pi$ (*k* is any integer), integrations of short periodic terms are zero. Selecting the minimum common periods of the long-term effects, integrations of long periodic terms are zero. Then there exists a minimum common interval over that the integration of all periodic terms are zero. The *d* for the first Keplerian element (*a*) is zero; i.e. the semi-major axis is not perturbed by the solar attracting force long periodically. The most important terms are the long-term effects which are represented in *L* and which are zero except the following two terms:

$$L_3 = \frac{\sqrt{1-e^2}}{n}\frac{3m\mu_s}{r_s^3}\left(\frac{-1,823}{16,384}n_1n_2 + \frac{9}{16}n_1^2 + \frac{9}{16}n_2^2 + \frac{3}{8}n_3^2\right)t$$
$$= \frac{\sqrt{1-e^2}}{n}\frac{3m\mu_s}{r_s^3}\left(\frac{-1,823}{16,384}n_1n_2 + \frac{9}{16} - \frac{3}{16}n_3^2\right)t \quad ,$$

$$L_6 = \frac{-3}{n}\frac{m\mu_s}{r_s^3}\left(\frac{21}{16}n_1^2 + \frac{21}{16}n_2^2 + \frac{7}{8}n_3^2\right)t = \frac{-3}{n}\frac{m\mu_s}{r_s^3}\left(\frac{21}{16} - \frac{7}{16}n_3^2\right)t. \qquad (93)$$

The results of the long-term perturbations are coincidently similar with that of the solutions of the part one given in (80); however, with different signs. Comparing (80) with (93) it is notable that the linear terms in (80) are generally (with a factor of $1/a$) smaller than that of (93). That is the linear effects are dominated by (93) under the solar attracting force perturbation. The orientation of the ellipse will rotate with a constant velocity in direction of increasing ω; whereas the mean anomaly (M) decreasing constantly.

Because the Gaussian equations are derived from the Lagrangian equations and there exist potential functions of the approximated disturbing force of the Sun, solutions via Gaussian equations must be the same and do not need to be discussed for multi-body disturbances.

4.4.5 Solutions of Disturbance of the Moon

The disturbance acceleration of the Moon is (see (69))

$$\vec{f}_m = -m\mu_m\left(\frac{1}{|\vec{r}-\vec{r}_m|^2}\vec{n}_{sm} + \frac{1}{|\vec{r}_m|^2}\vec{n}_m\right), \quad \vec{n}_{sm} = \frac{\vec{r}-\vec{r}_m}{|\vec{r}-\vec{r}_m|}, \quad \vec{n}_m = \frac{\vec{r}_m}{|\vec{r}_m|}. \qquad (94)$$

The unit vectors $\vec{n}_{\rm sm}$ and $\vec{n}_{\rm m}$ represent the vectors from the Moon to the satellite and the geocentric vector of the Moon, respectively. The only difference between (75) and (94) is the index; instead of "s" for the Sun, "m" is used for the Moon. However, the maximum geocentric distance rate of GPS satellite to the Moon is about 6×10^{-2}. This fact has to be taken into account and the discussions can be made similarly as did in Sect. 4.4.4. The force vector can be approximated by

$$\vec{f}_{\rm m} = -m\mu_{\rm m}\left(\frac{r}{r_{\rm m}^3}\left(1+3\frac{r}{r_{\rm m}}\cos\alpha\right)\vec{n}_{\rm r} - \frac{r}{r_{\rm m}^3}\left(3\cos\alpha + \frac{3r}{2r_{\rm m}}(5\cos^2\alpha - 1)\right)\vec{n}_{\rm m}\right). \tag{95}$$

The potential function of the disturbing force of the Sun is then

$$V_{\rm s} = -m\mu_{\rm m}\frac{r}{r_{\rm m}^3} + m\mu_{\rm m}\frac{3}{2r_{\rm m}^3}(\vec{n}_{\rm m}\cdot\vec{r})^2 - m\mu_{\rm m}\frac{1}{2r_{\rm m}^4}(-5r^3\cos^3\alpha + 3r^3\cos\alpha). \tag{96}$$

The solutions of Lagrangian equations related to the first two terms of potential function (96) are already derived in Sect. 4.4.4 and can be used directly. The solutions related to the last term can be derived similarly in principle and the discussions are omitted here (see Xu et al. 2010a).

Discretisation and Solution

Denote the satellite period as T. The local noon of the Moon is selected as the starting point of counting. A so-called sign function can be defined as

$$\delta(t) = \begin{cases} -1 & 0 \le t < T/2 \\ 1 & T/2 \le t \le T \end{cases}. \tag{97}$$

The sign function shows that the attracting force of the Moon decelerates the satellite during the first half period and accelerates during the second half period with respect to the nominal motion of the satellite. Then the duration of one period of $0 \sim T$ can be equally divided by Δt, i.e. by $t_0', t_1', \ldots, t_k', \ldots, T$. The acceleration of the disturbance of the Moon (94) is then discretised as

$$\vec{a}_{\rm m}(t) = -\mu_{\rm m}\left(\frac{1}{|\vec{r}(t_k) - \vec{r}_{\rm m}(t_k)|^2}\vec{n}_{\rm sm}(t_k) + \frac{1}{|\vec{r}_{\rm m}(t_k)|^2}\vec{n}_{\rm m}(t_k)\right).$$

The disturbed velocity caused by the Moon is then

$$\vec{v}_{\rm m}(t) = -\sum_{i=1}^{k}\mu_{\rm m}\left(\frac{1}{|\vec{r}(t_i) - \vec{r}_{\rm m}(t_i)|^2}\vec{n}_{\rm sm}(t_i) + \frac{1}{|\vec{r}_{\rm m}(t_i)|^2}\vec{n}_{\rm m}(t_i)\right)\Delta t.$$

The disturbed position caused by the Moon is then

$$\vec{\rho}_{\mathrm{m}}(t) = \sum_{j=1}^{k} \vec{v}_{\mathrm{m}}(t_j)\Delta t. \tag{98}$$

Equation (98) is the discrete solution of the disturbance of the Moon on the orbit of the satellite.

4.4.6 Solutions of Disturbance of Planets

The disturbance acceleration of a planet is (see (69))

$$\vec{f}_{\mathrm{p}} = -m\mu_{\mathrm{p}}\left(\frac{1}{\left|\vec{r}-\vec{r}_{\mathrm{p}}\right|^2}\vec{n}_{\mathrm{sp}} + \frac{1}{\left|\vec{r}_{\mathrm{p}}\right|^2}\vec{n}_{\mathrm{p}}\right), \quad \vec{n}_{\mathrm{sp}} = \frac{\vec{r}-\vec{r}_{\mathrm{p}}}{\left|\vec{r}-\vec{r}_{\mathrm{p}}\right|}, \quad \vec{n}_{\mathrm{p}} = \frac{\vec{r}_{\mathrm{p}}}{\left|\vec{r}_{\mathrm{p}}\right|}. \tag{99}$$

The unit vectors $\vec{n}_{\mathrm{sp}}$ and $\vec{n}_{\mathrm{p}}$ represent the vectors from the planet to the satellite and the geocentric vector of the planet, respectively. The geocentric distance of the planet is far greater than that of the Moon. The discussions in Sect. 4.4.5 can be directly used here. Because the disturbances of the planets are of the third order; the influences are need to be considered only in case of third-order solutions.

4.4.7 Summary

Solutions of the extraterrestrial disturbances of the attracting forces of the Sun and the Moon, as well as the planets, the drag force of the atmosphere and solar radiation pressure are derived in this section.

The solar radiation is a non-conservative disturbing force; of course, the disturbances of the orbit are also non-conservative ones. They are generally non-periodic effects.

The disturbance of the Sun has long-term effects on the orientation of the ellipse and the position of the satellite as well as short periodic effects on the semi-axis of the satellite and the shape of the ellipse. The effects of the Moon and planets are similar to that of the Sun.

4.5 Solutions of Geopotential Perturbations

The principle of the derivation of geopotential perturbations will be discussed first. The general solution of the perturbations of $\overline{C}_{lm}$ and $\overline{S}_{lm}$ is derived. Because of the length of the formulas, examples will be not given here. Interested reader may find them in Xu (2008) and on going publications.

Principle of the Derivations

From the solution process of the equation of satellite motion perturbed by the geopotential term $\overline{C}_{20}$ given in Chap. 5 of Xu (2008), one notices that the derivation is very complicated, even if the potential function of the perturbation is relatively simple. An alternative method is to use symbolic mathematical operation software such as Mathematica, Maple. However, the principle and strategy of the derivation have still to be carefully created.

For simplification, geopotential disturbance function of l order and m degree can be written as (see (4.35) of Xu 2008)

$$R_{lm} = \frac{\mu}{r}\left(\frac{a_e}{r}\right)^l \bar{P}_{lm}(\sin\varphi)\Big[\overline{C}_{lm}\cos\, m\lambda + \overline{S}_{lm}\sin\, m\lambda\Big]. \tag{100}$$

Let

$$\begin{aligned} \overline{C}_{lm} &= D_{lm}\cos\, m\lambda_{lm} \\ \overline{S}_{lm} &= D_{lm}\sin m\lambda_{lm}, \\ \overline{\lambda} &= \lambda - \lambda_{lm} \end{aligned} \tag{101}$$

where

$$\begin{aligned} D_{lm} &= \sqrt{\overline{C}_{lm}^2 + \overline{S}_{lm}^2} \\ \cos\, m\lambda_{lm} &= \frac{\overline{C}_{lm}}{D_{lm}}, \\ \sin m\lambda_{lm} &= \frac{\overline{S}_{lm}}{D_{lm}} \end{aligned} \tag{102}$$

then (100) is

$$R_{lm} = \frac{b_{lm}}{r^{l+1}}\overline{P}_{lm}(\sin\,\varphi)\cos(m\overline{\lambda}), \tag{103}$$

where $b_{lm} = \mu a_e^l D_{lm}$.

To transform the geographic coordinates into the Keplerian variables, the following relations are needed (see (5.2) of Xu 2008):

$$\begin{aligned} &\sin\,\varphi = \sin\, i\, \sin\, u, \\ &\overline{\lambda} = \alpha - \Theta - \lambda_{lm} = (\Omega - \Theta - \lambda_{lm}) + (\alpha - \Omega), \\ &\cos(\alpha - \Omega) = \frac{\cos\, u}{\cos\,\varphi}, \\ &\sin(\alpha - \Omega) = \frac{\sin\, u\cos\, i}{\cos\,\varphi}. \end{aligned} \tag{104}$$

Further more there are (see Wang et al. 1979)

$$\cos(my) = \sum_{j=0}^{[m/2]} (-1)^j \binom{m}{2j} (\cos y)^{m-2j} (\sin y)^{2j},$$

$$\sin(my) = \sum_{j=0}^{[(m-1)/2]} (-1)^j \binom{m}{2j+1} (\cos y)^{m-2j-1} (\sin y)^{2j+1}, \tag{105}$$

where $[z]$ is the integer part of z and the binomial form has the well-known expression of

$$\binom{m}{k} = \frac{m!}{k!(m-k)!}. \tag{106}$$

Let

$$\begin{array}{l} \overline{\Omega} = \Omega - \Theta - \overline{\lambda}_{\mathrm{lm}} \\ y = \alpha - \Omega \end{array}, \tag{107}$$

then

$$\begin{aligned} &\cos m\overline{\lambda} = \cos(m\overline{\Omega} + my) = \cos m\overline{\Omega} \cos my - \sin m\overline{\Omega} \sin my \\ &= \frac{1}{\cos^m \varphi} \sum_{j=0}^{[m/2]} (-1)^j \cos m\overline{\Omega} \binom{m}{2j} (\cos u)^{m-2j} (\sin u \cos i)^{2j} - \\ &\frac{1}{\cos^m \varphi} \sum_{j=0}^{[(m-1)/2]} (-1)^j \sin m\overline{\Omega} \binom{m}{2j+1} (\cos u)^{m-2j-1} (\sin u \cos i)^{2j+1}. \end{aligned}$$

Note that there is a factor of $\cos^m\varphi$ in the expression of $\bar{P}_{\mathrm{lm}}(\sin\varphi)$; therefore, let

$$\begin{aligned} q(\Omega, u, i) = &\sum_{j=0}^{[m/2]} (-1)^j \cos m\overline{\Omega} \binom{m}{2j} (\cos u)^{m-2j} (\sin u \cos i)^{2j} - \\ &\sum_{j=0}^{[(m-1)/2]} (-1)^j \sin m\overline{\Omega} \binom{m}{2j+1} (\cos u)^{m-2j-1} (\sin u \cos i)^{2j+1} \end{aligned} \tag{108}$$

$$\begin{aligned} Q_{lm}(x) &= \overline{P}_{lm}(x)/(1-x^2)^{m/2} \\ &= N_{lm} \sum_{k=0}^{K} T_{lmk} x^{l-m-2k}, \end{aligned} \tag{109}$$

where K is the integer part of $(l–m)/2$ and the factors are

$$N_{lm} = \sqrt{\frac{(l-m)!(2l+1)(2-\delta_{0m})}{(l+m)!}},$$

$$T_{lmk} = \frac{(-1)^k(2\,l-2\,k)}{2^l k!(l-k)!(l-m-2\,k)!}. \tag{110}$$

One has

$$R_{lm} = \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) q(\Omega, u, i), \tag{111}$$

and then

$$\frac{\partial R_{lm}}{\partial a} = \frac{\partial R_{lm}}{\partial r}\frac{\partial r}{\partial a} = \frac{-(l+1)}{a} R_{lm}\ ,$$

$$\frac{\partial R_{lm}}{\partial \Omega} = \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) \frac{\partial q(\Omega, u, i)}{\partial \Omega}\ ,$$

$$\frac{\partial R_{lm}}{\partial\, i} = \frac{b_{lm}}{r^{l+1}} \frac{\partial Q_{lm}(x)}{\partial x}\frac{\partial x}{\partial i} q(\Omega, u, i) + \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) \frac{\partial q(\Omega, u, i)}{\partial i},$$

$$\frac{\partial R_{lm}}{\partial \omega} = \frac{b_{lm}}{r^{l+1}} \frac{\partial Q_{lm}(x)}{\partial x}\frac{\partial x}{\partial u} q(\Omega, u, i) + \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) \frac{\partial q(\Omega, u, i)}{\partial u},$$

$$\begin{aligned}\frac{\partial R_{lm}}{\partial e} = {} & \frac{b_{lm}(-l-1)}{r^{l+2}}\frac{\partial r}{\partial e} Q_{lm}(x) q(\Omega, u, i) + \frac{b_{lm}}{r^{l+1}} \frac{\partial Q_{lm}(x)}{\partial x}\frac{\partial x}{\partial u}\frac{\partial u}{\partial e} q(\Omega, u, i) \\ & + \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) \frac{\partial q(\Omega, u, i)}{\partial u}\frac{\partial u}{\partial e}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial R_{lm}}{\partial\, M} = {} & \frac{b_{lm}(-l-1)}{r^{l+2}}\frac{\partial r}{\partial M} Q_{lm}(x) q(\Omega, u, i) + \frac{b_{lm}}{r^{l+1}} \frac{\partial Q_{lm}(x)}{\partial x}\frac{\partial x}{\partial u}\frac{\partial u}{\partial M} q(\Omega, u, i) \\ & + \frac{b_{lm}}{r^{l+1}} Q_{lm}(x) \frac{\partial q(\Omega, u, i)}{\partial u}\frac{\partial u}{\partial M},\end{aligned} \tag{112}$$

where

$$\begin{aligned}\frac{\partial q(\Omega, u, i)}{\partial \Omega} = {} & m \sum_{j=0}^{[m/2]} (-1)^j b_1(m, j) \sin m\overline{\Omega} (\cos\, u)^{m-2j} (\sin\, u\, \cos\, i)^{2j} \\ & - m \sum_{j=0}^{[(m-1)/2]} (-1)^j b_2(m, j) \cos m\overline{\Omega} (\cos\, u)^{m-2j-1} (\sin\, u\, \cos\, i)^{2j+1} \\ b_1(m, j) = {} & -\binom{m}{2j} \\ b_2(m, j) = {} & \binom{m}{2j+1},\end{aligned} \tag{113}$$

$$\frac{\partial q(\Omega,u,i)}{\partial u}=\sum_{j=0}^{[m/2]}(-1)^j\cos m\bar{\Omega}\begin{pmatrix}b_3(m,j)(\cos u)^{m-2j-1}(\sin u)^{2j+1}\\+b_4(m,j)(\cos u)^{m-2j+1}(\sin u)^{2j-1}\end{pmatrix}(\cos i)^{2j}$$
$$-\sum_{j=0}^{[(m-1)/2]}(-1)^j\sin m\bar{\Omega}\begin{pmatrix}b_5(m,j)(\cos u)^{m-2j-2}(\sin u)^{2j+2}\\+b_6(m,j)(\cos u)^{m-2j}(\sin u)^{2j}\end{pmatrix}(\cos i)^{2j+1}$$
$$b_3(m,j)=-\binom{m}{2j}(m-2j)$$
$$b_4(m,j)=\binom{m}{2j}2j$$
$$b_5(m,j)=-\binom{m}{2j+1}(m-2j-1)$$
$$b_6(m,j)=\binom{m}{2j+1}(2j+1), \tag{114}$$

$$\frac{\partial q(\Omega,u,i)}{\partial i}=$$
$$-\sum_{j=0}^{[m/2]}(-1)^j b_4(m,j)\cos m\overline{\Omega}(\cos u)^{m-2j}(\sin u)^{2j}(\cos i)^{2j-1}\sin i \tag{115}$$
$$-\sum_{j=0}^{[(m-1)/2]}(-1)^j b_6(m,j)\sin m\overline{\Omega}(\cos u)^{m-2j-1}(\sin u)^{2j+1}(\cos i)^{2j}\sin i$$

$$\frac{\partial Q_{lm}(x)}{\partial x}=N_{lm}\sum_{k=0}^{K}W_{lmk}x^{l-m-2k-1},\qquad W_{lmk}=T_{lmk}(l-m-2k) \tag{116}$$

$$x=\sin\varphi=\sin u\sin i, \tag{117}$$

$$\frac{\partial x}{\partial u}=\cos u\sin i, \tag{118}$$

$$\frac{\partial x}{\partial i}=\sin u\cos i, \tag{119}$$

$$\begin{aligned}\sin u&=\sin(f+\omega)=\sin f\cos\omega+\cos f\sin\omega\\ \cos u&=\cos(f+\omega)=\cos f\cos\omega-\sin f\sin\omega.\end{aligned} \tag{120}$$

Further more, there is

$$\frac{1}{r^N}=\frac{1}{a^N(1-e^2)^N}(1+e\cos f)^N=\frac{1}{a^N(1-e^2)^N}\sum_{n=0}^{N}\binom{N}{n}e^n\cos^n f, \tag{121}$$

$$\frac{\partial f}{\partial(e,M)}=\left(\frac{2+e\cos f}{1-e^2}\sin f,\left(\frac{a}{r}\right)^2\sqrt{1-e^2}\right), \tag{122}$$

$$\frac{\partial r}{\partial(a,e,\omega,i,\Omega,M)}=\left(\frac{r}{a},-a\cos f,0,0,0,\frac{ae}{\sqrt{1-e^2}}\sin f\right). \tag{123}$$

These derivations lead to simplified formulas for the perturbation function and are necessary and enough to transform the differential equations of motion into functions of Keplerian variables. They are used to derive the solutions of perturbations of geopotential function in order and degrees of 2×0, 2×1, 2×2, 3×0 and are the basis for deriving the general solution of the perturbation of l order and m degree.

4.6 Principle of Numerical Orbit Determination

Recalling the discussions made in Sect. 4.2, the perturbed orbit of the satellite is the solution (or integration)

$$\vec{X}(t) = \vec{X}(t_0) + \int_{t_0}^{t} \vec{F}\,\mathrm{d}t, \tag{124}$$

which can be obtained by integrating the differential state equation under the initial condition

$$\begin{cases} \dot{\vec{X}}(t) = \vec{F} \\ \vec{X}(t_0) = \vec{X}_0 \end{cases}, \tag{125}$$

where $\vec{X}(t)$ is the instantaneous state vector of the satellite, $\vec{X}(t_0)$ is the initial state vector at time t_0 (denoted by $\vec{X}_0$), $\vec{F}$ is a function of the state vector $\vec{X}(t)$ and time t and

$$\vec{X} = \begin{pmatrix} \vec{r} \\ \dot{\vec{r}} \end{pmatrix} \quad \text{and} \quad \vec{F} = \begin{pmatrix} \dot{\vec{r}} \\ \vec{f}/m \end{pmatrix},$$

where $\vec{f}$ is the summated force vector of all possible force vectors acting on the satellite, m is the mass of satellite, and $\vec{r}$, $\dot{\vec{r}}$ are the position and velocity vectors of the satellite.

If the initial state vector and the force vectors are precisely known, then the precise orbits can be computed through the integration in (124). Expanding the integration time t into the future, the so-called forecasted orbits can be obtained. Therefore, suitable numerical integration algorithms are needed (see Sect. 4.8.2 in Xu 2008).

In practice, the precise initiate state vector and force models, which are related to the approximate initial state vector and force models, have to be determined. These can be realised through suitable parameterisation of the models in the GPS observation equations and then the parameters can be solved by adjustment or filtering.

We generally denote both the range and range rate together by ρ; their partial derivatives with respect to the orbit state vector (see Xu 2007) have the forms

$$\frac{\partial \rho}{\partial \vec{r}}, \frac{\partial \rho}{\partial \dot{\vec{r}}} \quad \text{or} \quad \frac{\partial \rho}{\partial \vec{X}}.$$

Therefore, the orbit parameter-related parts in the linearised GPS observation equation are

$$\frac{\partial \rho}{\partial (\vec{r}, \dot{\vec{r}})} \frac{\partial (\vec{r}, \dot{\vec{r}})}{\partial \vec{y}} \Delta \vec{y}^{\mathrm{T}} \quad \text{or} \quad \frac{\partial \rho}{\partial \vec{X}} \frac{\partial \vec{X}}{\partial \vec{y}} \Delta \vec{y}^{\mathrm{T}}, \tag{126}$$

where

$$\vec{y} = (\vec{X}_0, \vec{Y}), \quad \Delta \vec{y}^T = (\Delta \vec{X}_0, \Delta \vec{Y})^T, \quad \frac{\partial \vec{X}}{\partial \vec{y}} = \frac{\partial \vec{X}}{\partial (\vec{X}_0, \vec{Y})},$$

where $\vec{X}$, $\vec{Y}$ are the state vector of satellite and the parameter vector of the force models, and index 0 denotes the related initial vectors of time t_0. $\vec{y}$ is the total unknown vector of the orbit determination problem, the related correction vector is $\Delta \vec{y} = \vec{y} - \vec{y}_0$ and $\Delta \vec{X}_0$ is the correction vector of the initial state vector. The partial derivative of $\vec{X}$ with respect to $\vec{y}$ is called transition matrix which has the dimension of $6 \times (6 + n)$, where n is the dimension of vector $\vec{Y}$. The partial derivatives of the equation of motion of the satellite (see (126)) with respect to the vector $\vec{y}$ are

$$\frac{\partial \dot{\vec{X}}(t)}{\partial \vec{y}} = \frac{\partial \vec{F}}{\partial \vec{y}} = \frac{\partial \vec{F}}{\partial \vec{X}} \frac{\partial \vec{X}}{\partial \vec{y}} + \left(\frac{\partial \vec{F}}{\partial \vec{y}} \right)^*, \tag{127}$$

where the superscript $*$ denotes the partial derivatives of $\vec{F}$ with respect to the explicit parameter vector $\vec{y}$ in $\vec{F}$ and

$$D(t) = \left(\frac{\partial \vec{F}}{\partial \vec{X}} \right) = \begin{pmatrix} 0_{3\times 3} & E_{3\times 3} \\ \frac{1}{m} \frac{\partial \vec{f}}{\partial \vec{r}} & \frac{1}{m} \frac{\partial \vec{f}}{\partial \dot{\vec{r}}} \end{pmatrix} = \begin{pmatrix} 0_{3\times 3} & E_{3\times 3} \\ A(t) & B(t) \end{pmatrix},$$

$$C(t) = \left(\frac{\partial \vec{F}}{\partial \vec{y}} \right)^* = \begin{pmatrix} 0_{3\times 6} & 0_{3\times n} \\ 0_{3\times 6} & \frac{1}{m} \frac{\partial \vec{f}}{\partial \vec{Y}} \end{pmatrix} = \begin{pmatrix} 0_{3\times (6+n)} \\ G(t) \end{pmatrix}, \tag{128}$$

where E is an identity matrix; the partial derivatives can be found in Xu (2008). Notable that the force parameters are not functions of t. Therefore the order of the differentiations can be exchanged. Denoting the transition matrix by $\Phi(t,t_0)$, then (128) turns out to be

$$\frac{\mathrm{d}\Phi(t, t_0)}{\mathrm{d}t} = D(t)\Phi(t, t_0) + C(t). \tag{129}$$

Equation (129) is called a differential equation of the transition matrix or variation equation (see, e.g., Montenbruck and Gill 2000). Denoting

$$\Phi(t, t_0) = \begin{pmatrix} \Psi(t, t_0) \\ \dot{\Psi}(t, t_0) \end{pmatrix}, \tag{130}$$

an alternate expression of (129) can be obtained by substituting (130) and (128) into (129):

$$\frac{\mathrm{d}^2\Psi(t, t_0)}{\mathrm{d}t^2} = A(t)\Psi(t, t_0) + B(t)\frac{\mathrm{d}\Psi(t, t_0)}{\mathrm{d}t} + G(t). \tag{131}$$

The initial value matrix is (initial state vector does not depend on force parameters):

$$\Phi(t_0, t_0) = (E_{6\times 6} \quad 0_{6\times n}). \tag{132}$$

That is, in the GPS observation equation, the transition matrix has to be obtained by solving the initial value problem of the variation equation (129) or (131). The problem is traditionally solved by integration. An algebraic solution is derived by Xu (2003).

Limitations of the Numerical Orbit Determination

Real-time ability of the numerical orbit determination is limited first by the adjustment or filtering algorithms used.

If the classic least squares adjustment algorithm is used to solve the parameters of the orbit determination problem, it is not possible to obtain the solution in real time because of the size and dimension of the equations. The equations of IGS orbit determination are formed and solved daily. It takes from less than 1 h to several hours to compute the results depending, of course, on the computer used. The so-called rapid IGS orbits are partly computed using 23 h past data and 1 h updated data. In general, the classic least squares adjustment algorithm is not suitable for real-time purpose.

Sequential least squares algorithm and Kalman filtering technique are partly developed for real-time applications. The sequential least squares algorithm is a special case of the Kalman filtering, therefore, the discussions will be focused on the filtering method. Kalman filtering solves the equations of every epoch or every epoch-block by taking into account the information from the past to obtain the results. In this way the problem can be solved epoch-wise or epoch-block-wise depending on the property of the problem. For equations of orbit determination the problem is not solvable (or singular) for a few epochs because of the dimension of the unknowns. The equations of orbit determination are generally solvable in half an hour (see Xu 2004) or longer. That is, the filtering technique and the property of the equations of orbit determination make the real-time application of the numerical orbit determination very difficult.

Furthermore, in numerical orbit determination, the numerical integration algorithms have to be used to integrate the orbits and to solve the variation equations. The numerical integrator usually has a so-called integrator length. The selection of the integrator length depends on the accuracy requirement and the physical properties of functions that will be integrated and therefore is not free of choice. Usually in IGS orbit determination, the integrator length is selected as 5 min. This also restricts the real-time application of the numerical orbit determination.

Because of the adjustment and filtering techniques and the use of the numerical integrator as well as the properties of the physical problem, numerical orbit determination is difficult to be in real time.

4.7 Principle of Analytic Orbit Determination

Orbit determination aims to determine the initial orbital elements (i.e. the initial state vector of the satellite) and the unknown model parameters. The technique of numerical orbit determination is developed in a situation that, on one hand, one needs the technique; however on the other hand, one does not have analytic solutions of the disturbed equations of satellite motion. The key difference between the numerical and the analytic orbit determination is that the orbits are represented in the former algorithm by differential equations and in the latter algorithm by analytic formulas.

The perturbed orbit of the satellite is the solution (or integration) (see discussions in Chaps. 5, 6 and 7 of Xu 2008)

$$\sigma_j(t) = \sigma_j(t_0) + (G_j(t) - G_j(t_0)) \quad \text{where} \quad G_j(t) - G_j(t_0) = \int_{t_0}^{t} F_j \, \mathrm{d}t, \tag{133}$$

where $G_j(t)$ are the infinite integrations of the right functions of the equations of motion and are given explicitly by analytic formulas. Equations (133) have been obtained by integrating the disturbed equations of motion

$$\begin{cases} \dot{\sigma}_j(t) = F_j \\ \sigma_j(t_0) = \sigma_{j_0} \end{cases}, \tag{134}$$

where $\sigma_j(t)$ is the j-th Keplerian element, $\sigma_j(t_0)$ is the related initial value at time t_0, F_j is the related right function of the differential equation and is a function of disturbing forces.

If the initial Keplerian elements and the force functions are precisely known, then the precise orbits can be computed by using (134). Computing for time t in the future, the so-called forecasted orbits can be obtained. That is, for orbit determination using analytic solutions, the traditional numerical integration algorithms are not necessary any more (because the differential equations are theoretically integrated by deriving the solutions).

In practice, the precise initial Keplerian elements are not known exactly and the parameters of the force models have to be co-determined. These can be realised through suitable parameterisation of the models in the GPS observation equations and then solved by adjustment or filtering.

We generally denote both the range and range rate together by ρ; their partial derivatives with respect to the orbit state vector (see Xu 2003, 2007) have the form

$$\frac{\partial \rho}{\partial \vec{r}}, \frac{\partial \rho}{\partial \dot{\vec{r}}}. \tag{135}$$

Therefore, the orbit parameter-related parts in the linearised GPS observation equation are then

$$\frac{\partial \rho}{\partial (\vec{r}, \dot{\vec{r}})} \frac{\partial (\vec{r}, \dot{\vec{r}})}{\partial (\sigma_j, j = 1, ..6)} \frac{\partial (\sigma_j, j = 1, \ldots 6)}{\partial \vec{y}} \Delta \vec{y}^{\mathrm{T}}, \tag{136}$$

where

$$\vec{y} = (\vec{\sigma}_0, \vec{Y}), \quad \Delta \vec{y}^{\mathrm{T}} = (\Delta \vec{\sigma}_0, \Delta \vec{Y})^{\mathrm{T}}. \tag{137}$$

$\vec{\sigma}$, $\vec{Y}$ are the Keplerian element vector and the parameter vector of the force models and index 0 denotes the related initial vectors of time t_0. $\vec{y}$ is the total unknown vector of the orbit determination problem, the related correction vector is $\Delta \vec{y} = \vec{y} - \vec{y}_0$ and $\Delta \vec{\sigma}_0$ is the correction vector of the initial Keplerian element vector. The partial derivatives of the satellite state vector with respect to the Keplerian element vector are known and can be found in Sect. 4.5.4 of Xu (2008). The partial derivative of the Keplerian element vector with respect to $\vec{y}$ is called the transition matrix which has the dimension of $6 \times (6 + n)$, where n is the dimension of vector $\vec{Y}$. Because of the analytic solutions of the disturbed equations of motion, the partial derivatives of the Keplerian elements with respect to the vector $\vec{y}$ are almost given by the solutions explicitly. That is to say, by analytic orbit determination, the transition matrix is represented by analytic formulas instead of the so-called variation equations in the numerical algorithm. The variation equation has disappeared from the orbit determination process, so the numerical integration algorithms traditionally used to solve the variation equation are not necessary any more.

Note that the orbit disturbances are mostly linear functions of the parameters of the force models. Therefore, the partial derivatives of Keplerian element vector with respect to parameter vector $\vec{y}$ of the force models are directly the coefficients of the related force parameters. No special derivations of the partial derivatives are needed.

Compared to numerical orbit determination (Chap. 8 of Xu 2008), in analytic orbit determination, no variation equations need to be solved; no numerical integration algorithms are necessary; no special orbit-related partial derivatives have to be derived. These significant advantages should lead to more efficient algorithms and more accurate orbit determination.

Real-Time Ability of Analytic Orbit Determination

Using the analytic orbit theory the observation equation of the orbit determination problem can be formed easily epoch-wise. The equations are solvable for an epoch-block. Taking past information into account, the solvable equations of an epoch-block can be formed and solved in real time. Taking the information before the solved epoch-block into account, Kalman filtering technique can be used to determine the orbit in real time. This is very significant for applications of satellite technology nowadays and should be further studied intensively.

Properties of Analytic Orbit Determination

Initial Time Selection

In numerical orbit determination, the initial time is a matter of free choice. For numerical integration, it really does not matter from which time point one starts to integrate. However, in analytic orbit solution, nearly half of the formulas are functions of initial time point (another half of the formulas are infinite integrations and functions of instantaneous time). In turn, the functions of the initial time point are in terms of sines and cosines. Of course, theoretically the initial time point of orbit determination can be freely selected. However, if the initial time point is selected at that point such that the sines or cosines of mean anomaly are zero, the intensity of the computations can be reduced by 25%. That is, a suitable initial time selection is very important for analytic orbit determination.

Using General Models for Second-Order Geopotential Disturbances

As shown in Chap. 6 of Xu (2008), the solutions of the second-order geopotential disturbances are very long. Theoretically, any order and any degree of the disturbances can be derived; however, to program all the formulas into software will definitely be a problem. For orbit determination the second-order geopotential disturbances are small terms and they can be dealt with like corrections to the initial and nominal orbit. For short periodic terms, the solutions are formed by a set of functions of

$$\{\sin nM,\ \cos nM,\ n = 1, \ldots, N\}, \tag{138}$$

where M is the mean anomaly of the orbit; n is an integer index and has a truncation number N.

Similarly, for the long periodic terms of the second-order geopotential disturbances, the solutions can be formed by the following sets of functions:

$$\{\sin n\omega,\ \cos n\omega,\ n = 1, \ldots, I\},$$

$$\{\sin n\Omega,\ \cos n\Omega,\ n = 1, \ldots, J\},$$

$$\{\sin(n\omega + m\Omega),\ \cos(n\omega + m\Omega),\ n, m = 1, \ldots, K\}, \tag{139}$$

where m is an integer index; I, J and K are truncation numbers.

The general models of the solutions of the second-order geopotential disturbances are then

$$\sum_{n=1}^{N} (A_n \cos nM + B_n \sin nM) + \sum_{n=1}^{I} (C_n \cos n\omega + D_n \sin n\omega)$$
$$+ \sum_{n=1}^{J} (E_n \cos n\Omega + F_n \sin n\Omega) + \sum_{n,m=1}^{K} (G_{nm} \cos(n\omega + m\Omega) + H_{nm} \sin(n\omega + m\Omega), \tag{140}$$

where coefficients $(A_n, B_n, C_n, D_n, E_n, F_n, G_{nm}, H_{nm})$ can be considered as unknown and should be co-determined by orbit determination. The truncation numbers of (I, J, K) are generally much smaller than N because of the long periodic properties and shall be suitably selected through practical experiments.

4.8 Summary and Discussions

Summary

Complete theory of the satellite orbit is described in a condensed way in this chapter. Perturbed equations of satellite motion are discussed first; then singularity-free and simplified equations are given. The solutions of extraterrestrial disturbances, such as solar radiation pressure, atmospheric drag and the disturbance of the Sun, the Moon and planets, are derived; then solutions of geopotential disturbances are discussed. Numerical and analytical orbit determinations are dealt with before the summary and discussions.

Discussions

Simplified Singularity-Free Equations of Motion

As seen above (Sect. 4.3), the singularity problem has been solved by using simplified and singularity-free equations. The simplified orbit problem is described using simplified coordinates. The geometric meanings of the variables remain the same. The use of the traditional and partly non-geometric sensed variable set of (a, $h = \sin i \cos \Omega$, $k = -\sin i \sin \Omega$, $\xi = e \cos(\omega+\Omega)$, $\eta = -e \sin(\omega+\Omega)$, $\lambda = M+\omega+\Omega$) is obviously not an ideal choice. One of the important reasons for using the canonical transformation to represent the orbit equations is that the canonical equations are also singularity-free. After the disturbed equations of motion (24) or

(25) are singularity-free, the advantages of the use of canonical equations have to be carefully re-evaluated.

Analytic Solution vs. Numeric Solution

Solutions of the extraterrestrial disturbances are sometimes given both in analytic and numerical form (see, e.g., Sects. 4.4.2 and 4.4.6). The formulas of the discrete solutions are very easy to use for computation; however, they do not have clear geometric explanations for the effects of the disturbances.

Potential Functions of the Sun, Moon and Planets

An approximation has been used in the derivation of the potential function of the disturbing force of the Sun. Similar means have been used for the Moon and can also be used for the planets. Therefore, the related solutions are derived under a precondition that the approximation is allowed.

Confusion of Non-conservative Force with Conservative Effect

Solar radiation is a non-conservative disturbing force. It is said that such a non-conservative force has a conservative effect. This is confusing and is shown in Fig. 4.7 with an example of solar disturbance on a GEO satellite. One of the possible reasons for such confusion may come from the adjustment model of the solar radiation used in the numerical orbit determination. The models used in traditional orbit determination (see (4.72) of Xu 2008) are periodic functions of the orbit. No matter what results are obtained from the adjustment, the results are periodic (or conservative). If the determined models are used to interpret the effects of the solar radiation, confusion is then the consequence. This shows that the parameterisation is very important and the parameterisation should be physically reasonable.

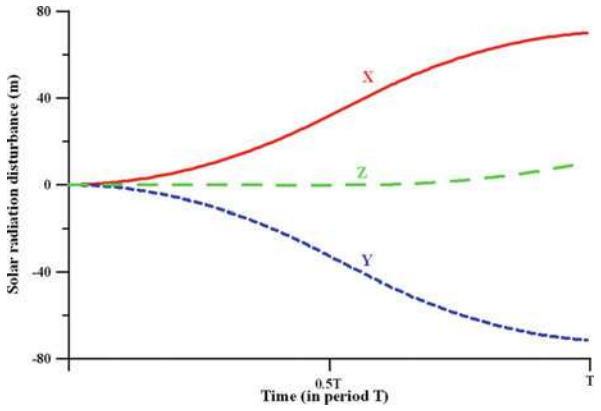


Fig. 4.7 Solar radiation disturbance on a GEO satellite

Long-Term Effects in Extraterrestrial Disturbances

There exist long-term effects in extraterrestrial disturbances (see (51)). The long-term perturbations have to be taken into account in the transformation of integral variables. This shall be particularly noticed in practical applications.

Long-Term and Long Periodic Effects in Short Periodic Disturbances

There exist long-term and long periodic effects in the short periodic geopotential disturbances (see (5.34) of Xu 2008). The long-term and long periodic effects derived in Sect. 4.5.2 of Xu (2008) are not unique and not the complete effects. Note that all the long-term and long periodic effects have to be accumulated if the relations are to be used.

Further Studies

Further studies have to be carried out on the analytic solutions of the Gaussian equations disturbed by the air drag, on the use of the simplified equations of motion, on the applications of the analytic theory (especially on the analytic orbit determination), on the study of the correlation of the geopotential disturbances on the orbits, and on the third-order solutions disturbed by the Earth and ocean tides as well as relativity disturbance.

Acknowledgements This study is under the frame of the internal GPS/Galileo study program of GFZ Potsdam and is partly supported by the Overseas Outstanding Chinese Scholar Fund of the Chinese Academy of Sciences hosted by the Institute of Geodesy and Geophysics in Wuhan.

References

Balmino, G., Schrama, E. and Sneeuw, N. (1996) Compatibility of first-order circular orbit perturbations theories: consequences for cross-track inclination functions. J. Geodesy, 70(9), 554–561

Bastos, L., Osorio, J. and Hein, G. (1995) GPS derived displacements in the Azores Triple Junction Region. In: GPS Trends in Precise Terrestrial Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1995. Springer, Berlin, pp. 99–104

Bate, R.R., Mueller, D.D. and White, J.E. (1971) Fundamentals of Astrodynamics. Dover Publications, New York

Battin, R.H. (1999) An Introduction to the Mathematics and Methods of Astrodynamics, revised version, AIAA Education Series. Massachusetts Institute of Technology, Cambridge

Beutler, G. (1994) GPS Trends in Precise Terrestrial, Airborne, and Space Borne Applications. Springer, Heidelberg

Beutler, G (1996) GPS satellite orbits. In: Kleusberg, A. and Teunissen, P.J.G. (eds) GPS for Geodesy. Springer, Berlin

Beutler, G., Brockmann, E., Gurtner, W., Hugentobler, U., Mervart, L., Rothacher, M. and Verdun, A. (1994) Extended orbit modelling techniques at the CODE Processing Center of the IGS: Theory and initial results. Manuscr Geodaet, 19, 367–386

Beutler, G., Brockmann, E. and Hugentobler, U. (1996) Combining consecutive short arcs into long arcs for precise and efficient GPS orbit determination. J. Geodesy, 70(5), 287–299

Beutler, G., Schildknecht, T., Hugentobler, U. and Gurtner, W. (2003) Orbit determination in satellite geodesy. Adv. Space Res., 31(8), 1853–1868

Boomkamp, H. and Dow, J. (2005) Use of double difference observations in combined orbit solutions for LEO and GPS satellites. Adv. Space Res. 36(3), 382–391

Borre, K. (2003) The GPS Easy Suit-Matlab code for the GPS newcomer. GPS Solutions, 7(1), 47–51

Bronstein, I.N. and Semendjajew, K.A. (1987) Taschenbuch der Mathematik. B.G. Teubner Verlagsgesellschaft, Leipzig, Germany. ISBN 3-322-00259-4

Brouwer, D. and Clemence, G.M. (1961) Methods of Celestial Mechanics. Academic Press, New York

Casotto, S. and Zin, A. (2000) An assessment of the benefits of including GLONASS data in GPS-based precise orbit determination – I: S/A analysis. Adv. Astronaut. Sci., 105(1), 237–256

Chobotov, V.A. (ed) (1991) Orbital Mechanics. AIAA, Washington

Colombo, O.L. (1984a) Altimetry, orbits and tides. NASA Technical Memorandum 86180

Colombo, O.L. (1984b) The global mapping of gravity with two satellites. Netherlands Geodetic Commission, Delft, The Netherlands Publications on Geodesy, 7(3), 253 pp

Cui, C. (1990) Die Bewegung künstlicher Satelliten im anisotropen Gravitationsfeld einer gleichmässig rotierenden starren Modellerde. Deutsche Geodätische Kommission, Reihe C: Dissertationen, Heft Nr. 357

Cui, C. (1997) Satellite orbit integration based on canonical transformations with special regard to the resonance and coupling effects. Dtsch Geod Komm bayer Akad Wiss, Reihe A, Nr. 112, 128 pp

Cui, C. and Lelgemann, D. (1995) Analytical dynamic orbit improvement for the evaluation of geodetic-geodynamic satellite data. J. Geodesy, 70, 83–97

Davis, P. and Rabinowitz, P. (1984) Methods of Numerical Integration, 2nd Ed. Academic Press, New York

Desai, S.D. and Haines, B.J. (2003) Near-real-time GPS-based orbit determination and sea surface height observations from the Jason-1 mission. Mar. Geod. 26(3/4), 383–397

Dick, G. (1997) Nutzung von GPS zur Bahnbestimmung niedrigfliegender Satelliten. GPS-Anwendungen und Ergebnisse '96: Beiträge zum 41. DVW-Fortbildungsseminar vom 7. bis 8. November 1996 am Geo-Forschungszentrum Potsdam, pp. 241–249

Douša, J. (2004) Precise orbits for ground-based GPS meteorology: processing strategy and quality assessment of the orbits determined at geodetic observatory. Pecny. J. Meteor. Soc. Jpn., 82(1B), 371–380

Dow, J.M. (1988) Ocean tides and tectonic plate motions from Lageos. Deutsche Geodätische Kommission, Rheihe C, Dissertation, Heft Nr. 344

Dow, J.M., Romay-Merino, M.M., Piriz, R., Boomkamp, H. and Zandbergen, R. (1993) High precision orbits for ERS-1: 3-day and 35-day repeat cycles. Proceedings of the 2nd ERS-1 Symposium – Space at the Service of our Environment, Hamburg, 11–14 October 1993, pp. 1349–1354

Exertier, P. and Bonnefond, P. (1997) Analytical solution of perturbed circular motion: Application to satellite geodesy. J. Geodesy, 71(3), 149–159

Feltens, J. (1991) Nicht gravitative Störeinflüsse bei der Modellierungen von GPS-Erdumlaufbahnen. Verlag der Bayerischen Akademie der Wissenschaften, Munich, Deutsche Geodätische Kommision, Reihe C, Heft Nr. 371

Gabor, M.J. and Nerem, R.S. (2004) Characteristics of satellite-satellite single difference widelane fractional carrier-phase biases. Navigation, 51(1), 77–92

Goad, C., Dorota, A., Brzezinska, G. and Yang, M. (1996) Determination of high-precision GPS orbits using triple differencing technique. J. Geodesy, 70, 655–662

Guinn, J., Muellerschoen, R. and Cangahuala, L. (1995) TOPEX/Poseidon precision orbit determination using combined GPS, SLR and DORIS. In: GPS Trends in Precise Terrestrial,

Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1995, Springer, Berlin, pp. 128–132
Haines, B.J., Christensen, E.J. and Guinn, J.R. (1995) Observations of TOPEX/Poseidon orbit errors due to gravitational and tidal modeling errors using the Global Positioning System. In: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1995. Springer, Berlin, pp. 133–138
Haines, B., Bar-Server, Y., Bertiger, W., Desai, S. and Willis, P. (2004) One-centimeter orbit determination for Jason-1: New GPS-based strategies. Mar. Geod., 27(1–2), 299–318
Hatch, R.R. (2004) Those scandalous clocks. GPS Solutions, 8(2), 67–73
Herrick, S. (1972) Astrodynamics, Vol. II. Van Nostrand Reinhold, London
Herring, T. (2003) MATLAB Tools for viewing GPS velocities and time series. GPS Solutions, 7(3), 194–199
Hess, D. and Keller, W. (1999a) Gradiometrie mit GRACE, Teil I: Fehleranalyse künstlicher Gradiometerdaten. ZfV, 5, 137–144
Hess, D. and Keller, W. (1999b) Gradiometrie mit GRACE, Teil II: Simulationsstudie. ZfV, 7, 205–211
Hofmann-Wellenhof, B., Lichtenegger, H. and Collins, J. (1997, 2001) GPS Theory and Practice. Springer-Press, Wien, Austria
Hugentobler, U., Schaer, S. and Fridez, P. (2001) Bernese GPS Software (Version 4.2). Astronomical Institute of University, Berne
Hugentobler, U., Ineichen, D. and Beutler, G. (2003) GPS satellites: Radiation pressure, attitude and resonance. Adv. Space Res., 31(8), 1917–1926
Jaggi, A., Beutler, G. and Hugentobler, U. (2005) Reduced-dynamic orbit determination and the use of accelerometer data. Adv. Space Res., 36(3), 438–444
Kang, Z. (1998) Präzise Bahnbestimmung niedrigfliegender Satelliten mittels GPS und die Nutzung für die globale Schwerefeldmodellierung. Scientific Technical Report STR 98/25, GeoForschungsZentrum (GFZ) Potsdam
Kang, Z., Nagel, P. and Pastor, R. (2003) Precise orbit determination for GRACE. Adv. Space Res., 31(8), 1875–1881
Karslioglu, M.O. (2005) An interactive program for GPS-based dynamic orbit determination of small satellites. Comput. Geosci., 31(3), 309–317
Kaula, W.M. (1966/2001) Theory of Satellite Geodesy. Blaisdell Publishing Company, Dover Publications, New York
Kleusberg, A. (1995) Mathematics of attitude determination with GPS. GPS World, 6(9), 72–78
Konig, R., Reigber, C. and Zhu, S.Y. (2005) Dynamic model orbits and earth system parameters from combined GPS and LEO data. Adv. Space Res., 36(3), 431–437
Kroes, R. and Montenbruck, O. (2004) Spacecraft formation flying: Relative positioning using dual-frequency carrier phase. GPS World, 15(7), 37–42
Kroes, R., Montenbruck, O., Bertiger, W. and Visser, P. (2005) Precise GRACE baseline determination using GPS. GPS Solutions, 9(1), 21–31
Kuang, D., Rim, H.J., Schutz, B.E. and Abusali, P.A.M. (1996) Modeling GPS satellite attitude variation for precise orbit determination. J. Geodesy, 70, 572–580
Kwon, J.H., Grejner-Brzezinska, D., Bae, T.S. and Hong, C.K. (2003) A triple difference approach to Low Earth Orbiter precision orbit determination. J. Navig., 56(3), 457–473
Lelgemann, D. (1983) A linear solution of equation of motion of an Earth-orbiting satellite based on a Lie-series. Celestial Mech., 30, 309
Lelgemann, D. (1996) Geodaesie im Weltraumzeitalter. Dtsch Geod Komm, 25, 59–77
Lelgemann, D. (2002) Lecture Notes of Geodesy. Technical University, Berlin
Lemmens R (2004) Book review: GPS – theory, algorithms and applications, Xu G 2003. Int. J. Appl. Earth Obs. Geoinf., 5, 165–166
Liu, D.J., Shi, Y.M. and Guo, J.J. (1996) Principle of GPS and its Data Processing. TongJi University Press, Shanghai (in Chinese)

Liu, L. and Zhao, D. (1979) Orbit Theory of the Earth Satellite. Nanjing University Press, Nanjing (in Chinese)

Mackenzie, R. and Moore, P. (1997) A geopotential error analysis for a non planar satellite to satellite tracking mission. J Geodesy, 71(5), 262–272

Mackie, J.B. (1985) The Elements of Astronomy for Surveyors. Charles Griffin & Company Ltd., London

Mansfeld, W. (2004) Satellitenortung und Navigation, 2nd Ed. Vieweg Verlag, Wiesbanden, 352 pp.

McCarthy, D.D. (1996) International earth rotation service. In: IERS conventions, Paris, 95 pp. IERS Technical Note No. 21

McCarthy, D.D. and Capitaine, N. (2002) Practical Consequences of Resolution B1.6 "IAU2000 Precession-Nutation Model", Resolution B1.7 "Definition of Celestial Intermediate Pole", and Resolution B1.8 "Definition and Use of Celestrial and Terrestrial Ephemeris Origin". In: Capitaine, N., et al. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, Paris, April 18–19, 2002. IERS Technical Note No. 29

McCarthy, D.D. and Luzum, B.J. (1995) Using GPS to determine Earth orientation. In: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1999. Springer, Berlin, pp. 52–58

McCarthy, D.D. and Petit, G. (eds) (2003) International earth rotation service. IERS conventions (2003), IERS Technical Note No. 32

Meeus, J. (1992) Astronomische Algorithmen. Johann Ambrosius Barth, Verlag

Melchior, P. (1978) The Tides of the Planet Earth. Pergamon Press, Brussels

Mireault, Y., Kouba, J. and Lahaye, F. (1995) IGS combination of precise GPS satellite ephemerides and clock. In: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1995. Springer, Berlin, pp. 14–23

Mitchell, S., Jackson, B. and Cubbedge, S. (1996) Navigation solution accuracy from a spaceborne GPS receiver. GPS World, 7(6), 42, 44, 46–48, 50

Montenbruck, O. (1989) Practical Ephemeris Calculations. Springer, Heidelberg

Montenbruck, O. (2003) Kinematic GPS positioning of LEO satellites using ionospheric-free single frequency measurements. Aerosp. Sci. Technol., 7(5), 396–405

Montenbruck, O. and Gill, E. (2000) Satellite Orbits: Models, Methods and Applications. Springer, Heidelberg

Montenbruck, O. and Kroes, R. (2003) In flight performance analysis of the CHAMP BlackJack GPS receiver. GPS Solutions, 7(2), 74–86

Montenbruck, O., Gill, E. and Kroes, R. (2005) Rapid orbit determination of LEO satellites using IGS clock and ephemeris products. GPS Solutions, 9(3), 226–235

Montenbruck, O., van Helleputte, T., Kroes, R. and Gill, E. (2005) Reduced dynamic orbit determination using GPS code and carrier measurements. Aerosp. Sci. Technol., 9(3), 261–271

Moritz, H. (1980) Advanced Physical Geodesy. Herbert Wichmann Verlag, Karlsruhe

Mueller, I.I. (1964) Introduction to Satellite Geodesy. Frederick Ungar Publishing, New York

Murakami, M. (1996) Precise determination of the GPS satellite orbits and its new applications: GPS orbit determination at the Geographical Survey Institute. J. Geod. Soc. Jpn., 42(1), 1–14

Parkinson, B.W. and Spilker, J.J. (eds) (1996) Global Positioning System: Theory and Applications. 2 Vols, Progress in Astronautics and Aeronautics, Vol. 163, Washington

Petit, G. (2002) Comparison of "Old" and "New" Concepts: Coordinate Times and Time Transformations. In: Capitaine, N., et al. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, Paris, April 18–19, 2002. IERS Technical Note No. 29

Rapp, R.H. (1986) Global geopotential solutions. In: Sunkel, H. (ed) Mathematical and Numerical Techniques in Physical Geodesy. Lecture Notes in Earth Sciences, Vol. 7. Springer, Heidelberg.

Reigber, C. (1997) Geowissenschaftlicher Kleinsatellit CHAMP. GPS-Anwendungen und Ergebnisse '96: Beiträge zum 41. DVW-Fortbildungsseminar vom 7. bis 8. November 1996 am Geo-Forschungs-zentrum Potsdam, pp. 266–273

Remondi, B. (1984) Using the Global Positioning System (GPS) phase observable for relative geodesy: Modelling, processing, and results. PhD Dissertation, Center for Space Research, University of Texas at Austin, Austin

Roßbach, U. (2006) Positioning and Navigation Using the Russian Satellite System GLONASS, Universität der Bundeswehr München, URN: de:bvb, pp. 707–648

Rothacher, M. and Mervart, L. (1996) Bernese GPS Software (Version 4.0). Astronomical Institute of University of Bern, Bern

Rothacher, M. and Schaer, S. (1995) GPS-Auswertetechniken. Schriftenreihe des Deutschen Vereins für Vermessungswesen, Bd. 18, Stúttgart, pp. 107–121

Rummel, R. and Van Gelderen, M. (1995) Meissl scheme: Spectral characteristics of physical geodesy. Manuscr Geodaet, 20(5), 379–385

Rush, J. (2000) Current issues in the use of the global positioning system aboard satellites. Acta Astronaut., 47(2–9), 377–387

Scheinert, M. (1996) Zur Bahndynamik niedrigfliegender Satelliten. Verlag der Bayerischen Akademie der Wissenschaften, DGK, Reihe C, Heft 435 .

Schneider, M. (1988) Satellitengeodaesie. Wissenschaftsverlag, Mannheim

Schneider, M. and Cui, C.F. (2005) Theoreme über Bewegungsintegrale und ihre Anwendung in Bahntheorien, Bayerischen Akad Wiss, Reihe A, Heft Nr. 121, München, 132 pp

Schutz, B.E. (2000) Numerical studies in the vicinity of GPS deep resonance. Adv. Astronaut. Sci. 105(1), 287–302

Schwintzer, P., Kang, Z. and Reigber, C. (1995) GPS satellite-to-satellite tracking for TOPEX/Poseidon precise orbit determination and gravity field model improvement. J. Geodyn., 20(2), 155–166

Seeber, G. (1989, 1993) Satelliten-Geodaesie. Walter de Gruyter, Berlin, New York

Seeber, G. (1996) Grundprinzipien zur Vermessung mit GPS. Vermessungsingenieur, 47(2), 53–64

Seeber, G. (2003) Satellite Geodesy: Foundations, Methods, and Applications. Walter de Gruyter, Berlin, 589 pp

Sigl, R. (1978) Geodätische Astronomie. Wichmann Verlag, Karlsruhe

Sigl, R. (1989) Einführung in die Potentialtheorie. Wichmann Verlag, Karlsruhe

Smith, A.J.E., Hesper, E.T., Kuijper, D.C., Mets, G.J., Visser, P.N., Ambrosius, B.A.C. and Wakker, K.F. (1996) TOPEX/ Poseidon orbit error assessment. J. Geodesy, 70, 546–553

Springer, T.A., Beutler, G. and Rothacher, M. (1999) Improving the orbit estimates of GPS satellites. J. Geodesy, 73, 147–157

Strang, G. and Borre, K. (1997) Linear Algebra, Geodesy, and GPS. Wellesley-Cambridge Press, Wellesley

Torge, W. (1991) Geodesy. Walter de Gruyter, Berlin

Van Kamp, P.D. (1967) Principles of Astronomy. W.H. Freemann and Company, San Francisco, CA/London

Visser, P.N.A.M. and Van den Ijssel, J. (2000) GPS-based precise orbit determination of the very low Earth-orbiting gravity mission GOCE. J. Geodesy, 74(7/8), 590–602

Wang, L.X., Fang, Z.D., Zhang, M.Y., Lin, G.B., Gu, L.K., Zhong, T.D., Yang, X.A., She, D.P., Luo, Z.H., Xiao, B.Q., Chai, H. and Lin, D.X. (1979) Mathematic Handbook. Educational Press, Peking, ISBN 13012-0165

Weber, R. (1996) Monitoring Earth orientation variations at the Center for Orbit Determination in Europe (CODE). Oesterr Z Vermess Geoinf, 84(3), 269–275

Wenzel, H.-G. (1985) Hochauflösende Kugelfunktionsmodelle für das Gravitationspotential der Erde. Wissenschaftliche Arbeiten der TU Hannover, Nr. 137, Hannover

Xu, G. (2003) GPS - Theory, Algorithms and Applications. Springer, Heidelberg, 315 pp

Xu, G. (2004) MFGsoft – Multi-Functional GPS/(Galileo) Software: Software User Manual, (Version of 2004), Scientific Technical Report STR04/17 of GeoForschungsZentrum (GFZ) Potsdam, ISSN 1610-0956, 70 pp. www.gfz-potsdam.de/bib/pub/str0417/0417.pdf

Xu, G. (2007) GPS – Theory, Algorithms and Applications, 2nd Ed. Springer, Heidelberg, 350 pp

Xu, G. (2008) Orbits. Springer, Heidelberg, 230 pp

Xu, G., Schwintzer, P. and Reigber, Ch. (1998) KSGSoft (Kinematic/Static GPS Software): Software User Manual (Version of 1998). Scientific Technical Report STR98/19 of GeoForschungsZentrum (GFZ) Potsdam

Xu, G., Xu, T., Chen, W., and Yeh, T-K. (2010a) Analytic Solution of Satellite Orbit Disturbed by Lunar and Solar Gravitation, MNRAS (accepted for publication)

Xu, G., Xu, T., Yeh, T-K. and Chen W. (2010b) Analytic Solution of Satellite Orbit Disturbed by atmospheric Drag, MNRAS (accepted for publication)

Xu, Q.F. (1994) GPS Navigation and Precise Positioning. Army Press, Peking, ISBN 7-5065-0855-9/P.4 (in Chinese)

Yang, Y. (1997) Robust Kalman filter for dynamic systems. J. Zhengzhou. Inst. Surv. Mapp., 14, 79–84

Yang, Y. and Cui, X. (2008) Adaptively Robust Filter with Multi Adaptive Factors. Survey Review, 40(309), 260–270

Yang, Y., He, H. and Xu, G. (2001) Adaptively robust filtering for kinematic geodetic positioning. J. Geodesy, 75, 109–116

Yeh, T.K. and Chen, C.S. (2006) Clarifying the relationship between the clock errors and positioning precision of GPS receiver. In: VI Hotline-Marussi Symposium of Theoretical and Computational Geodesy Wuhan, Wuhan.

Yoon, J.C., Lee, B.S. and Choi, K.H. (2000) Spacecraft orbit determination using GPS navigation solutions. Aerosp. Sci. Technol., 4(3), 215–221

Yuan, Y.B. and Ou, J.K. (1999) The effects of instrumental bias in GPS observations on determining ionospheric delays and the methods of its calibration. Acta Geod Cartogr Sinica 28(2)

Yunck, T.P. and Melbourne, W.G. (1995) Spaceborne GPS for earth science. In: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications: 21st IUGG General Assembly, IAG Symposium No. 115, Boulder, July 3–4, 1995. Springer, Berlin, pp. 113–122

Zhang, Q., Moore, P., Hanley, J. and Martin, S. (2006) Auto-BAHN: software for near real-time GPS orbit and clock computation. Adv. Space Res., 39(10), 1531–1538

Zhou, J., Huang, Y., Yang, Y. and Ou, J. (1997) Robust least squares method. Publishing House of Huazhong University of Science and Technology, Wuhan

Zhu, S. (1997) GPS-Bahnfehler und ihre Auswirkung auf die Positionierung. GPS-Anwendungen und Ergebnisse '96: Beiträge zum 41. DVW-Fortbildungsseminar vom 7. bis 8. November 1996 am GeoForschungszentrum Potsdam, pp. 219–226

Chapter 5
Deformation and Tectonics: Contribution of GPS Measurements to Plate Tectonics – Overview and Recent Developments

Luisa Bastos, Machiel Bos and Rui Manuel Fernandes

Contents

5.1 Introduction

The use of space-geodetic techniques to study geodynamic processes began with Very Long Baseline Interferometry (VLBI) in the early 1970s. By measuring the delay in arrival time of the signal from distant celestial objects, the distances between stations that are hundreds of kilometres apart can be derived with millimetre accuracy. A review of the first 20 years of this technique is given by Ryan and Ma (1998). Around the world there are nowadays more than 100 VLBI stations. Another technique that has been available since the early 1970s is Satellite Laser Ranging (SLR). As the name implies, this technique determines the distance to a satellite by measuring the round trip time of a light pulse that is sent to the satellite (Degnan 1993). Today, there are about 60 SLR stations operational around the world.

SLR and VLBI provided a fundamental contribution to clarify the kinematics of the tectonic plates on a global scale. During the 1980s, a quite extensive use of

L. Bastos (✉)
Observatório Astronómico, Faculdade de Ciências, Universidade do Porto, Alameda do Monte da Virgem, 4420-146 V. N. Gaia, Portugal
e-mail: lcbastos@fc.up.pt

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_5,

mobile VLBI and SLR systems was made in dedicated campaigns to monitor some specific regions of the world with more significant tectonic activity (Vermaat et al. 1998). In addition VLBI technique was, and still is, an important tool to support the realisation of an accurate and stable celestial reference system. SLR has an important role in the determination of the Earth's geocentre and in calibrating geodetic microwave techniques. The limitations for a wider use of VLBI and SLR for geodynamic purposes are their relative lack of portability, high cost and operational complexity.

The quick development of satellite technology has led to the rise, in the early 1980s, of Global Navigation Satellite Systems (GNSS), such as GPS (Global Positioning System) and GLONASS (GLObal NAvigation Satellite System). GPS rapidly started to be applied for geodynamic applications, whereas GLONASS, due to difficulties in achieving its full operational implementation, has not yet been used in many geodynamic studies.

In the 1990s, radio satellite systems like DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite) started to give useful information for the study of regional and global geodynamics issues. Soudarin and Grétaux (2006) present a recent geodynamic study using 57 DORIS stations.

Actually, the most commonly used space-geodetic technique for geodynamic studies is still GPS. A thorough description of the GPS technique can be found in Hoffman-Wellenhof et al. (1997) and Parkinson et al. (1996).

To achieve the very high accuracy needed for deformation studies, one needs to look at a wide range of aspects related to GPS processing. Examples are satellite dynamics, reference frame definition, Earth orientation, ionospheric and tropospheric delay corrections, ambiguity resolution on a global scale, tidal and other loading effects, multipath and antenna phase centre variation. A review of the use of GPS for geodetic applications was written 10 years ago by Herring (1999). Other review works related to geodynamic applications can be found in Blewitt (1993) and Segall and Davis (1997). Here we extend such reviews to the current day, putting the emphasis on the use of GPS to observe plate tectonic motions.

Until the early 1990s basically only campaign data were used to monitor the Earth's surface deformation and centimetre accuracy was foreseen. It was already then demonstrated that the goal of measuring horizontal deformations with an accuracy of 5 mm/year should be soon achieved. Repeatability of the observed position was a measure of accuracy of the GPS technique.

The continuous developments in GPS receiver technology, and a drop in equipment cost, have led to a rapid change in the observation methodology. In the early 1990s the campaign type approach started to be replaced by the permanent network approach with great impact on the study of the dynamics of the Earth. Examples of permanent networks that started to be deployed in the early 1990s are the Permanent GPS Geodetic Array (PGGA) in California (Bock et al. 1993) and a nationwide network established in Japan with hundreds of GPS stations (Tsuji et al. 1995; Miyazaki et al. 1996). Shimada and Bock (1992) reported some of the earliest results from a permanent GPS network.

The most well-known permanent network of GPS stations is operated by the IGS (International GPS Service, today called International GNSS Service) (Beutler

1996), with a worldwide coverage, which has a major role in defining the reference frame for all different applications of GNSS. The 1992 IGS Test Campaign served as the proof of concept for IGS. Two main goals led to the establishment of IGS: the need from the geodetic community for precise orbit determination and the need for other products such as Earth orientation parameters and a reference frame.

There are now 336 active permanent (GPS and GPS+GLONASS) stations contributing to the IGS. In Europe the EUREF permanent network (EPN) consists of around 190 continuously operating GPS reference stations (EPNCB 2008). Figure 5.1 shows the present-day distribution of permanent GPS, SLR, VLBI and DORIS stations that contribute to the realisation of the International Terrestrial Reference Frame.

Permanent networks reveal their usefulness to detect deformations between plates (inter-plate) and deformations inside the tectonic plate (intra-plate) as well. With continuous GPS data it is possible to study a wide range of transient deformation processes associated with seismotectonic activity and other geophysical signals that were not possible with episodic GPS field campaigns. Daily solutions from permanent stations yield more accurate estimates of average deformation rates than those derived from periodic campaign measurements. In Sect. 5.2 we discuss how GPS is contributing to clarify these aspects.

Absolute plate model velocities are given with respect to a global reference frame. The realisation of such a frame is not trivial. The current method is to use a global set of geodetic stations that are on stable plates, far away from the boundaries, as the realisation of the reference frame. Each station of this set has been given a position and velocity. In Sect. 5.3 we explain how a new geodetic station with unknown position and velocity can be mapped into this reference frame.

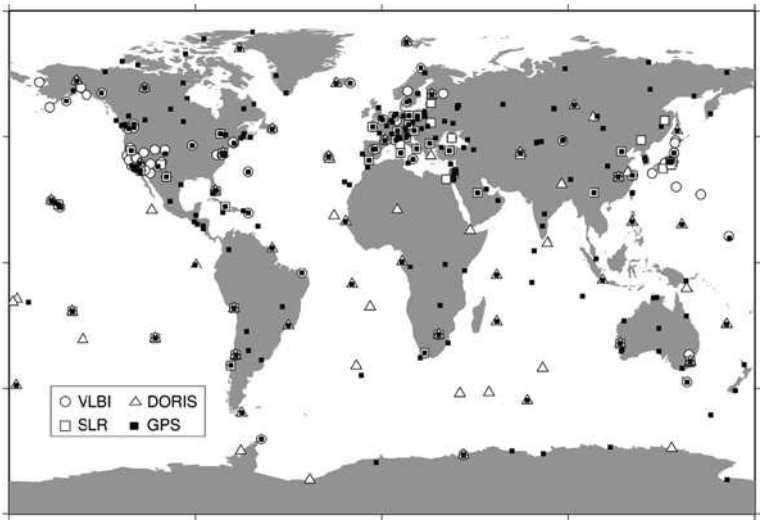


Fig. 5.1 Present-day distribution of VLBI, SLR, DORIS and GPS stations that contribute to the realisation of the international reference frame

An assumption that is always implicitly made is that geodetic stations represent the velocity of the tectonic plate on which they are located. Now that the accuracy has reached the millimetre level, one has to correct for many geophysical phenomena, which are mainly periodic, before this is true. In addition to periodic motions that may affect a station position, there are deformations due to the elastic response of the Earth to changes of atmospheric pressure, the nearly secular motions associated with post-glacial rebound and the occasional episodic motions associated with earthquakes. The degree of accuracy achieved in the determination of a site motion demands careful analysis of all different aspects involved in the GPS process, including a deep understanding of all the underlying geophysical phenomena. The geophysical corrections involved in the interpretation of a GPS time-series are described in detail in Sect. 5.4.

As mentioned before, the GPS observing strategy has evolved from the so-called campaign mode to the permanent network approach. In spite of the good accuracy achieved on the basis of epoch-type campaigns, this approach does not allow the discrimination between continuous deformations and instantaneous displacements associated with seismic or volcanic activity. Local or regional networks of permanently operating systems are more effective. This has also allowed for better investigation of the noise properties within the GPS time-series. As a result, also the way of estimating the tectonic motion from the GPS position time-series has evolved during the last decade. A summary of this history is given in Sect. 5.5.

A few examples of the impact of GPS-derived information in the study of the kinematics of the Azores region are presented in Sect. 5.6. Finally, future prospects in the exploitation of space-geodetic techniques to study the Earth's deformation are referred to.

5.2 Plate Tectonic Models

Alfred Wegener suggested at the beginning of the last century that the continents were once one large land mass that has undergone processes that gave rise to several tectonic plates, which drifted to their current locations. This theory could only be confirmed by the mid-1950s, when paleomagnetic data became available.

By the 1970s, several plate motion models, based on geophysical and geological data, were presented (Chase 1978; Minster and Jordan 1978) and these started to be adopted by the scientific community in general. These early models and, more recently, NUVEL-1 (DeMets et al. 1990) and its updated version, NUVEL-1A (DeMets et al. 1994), were computed using geological and geophysical data, such as ocean floor magnetic anomalies, transform faults and earthquake slip vectors, averaging over a period of 3–5 million years.

When published, NUVEL-1A was considered to provide the best estimate of the angular velocity for major plates and it is still used intensively these days. However, this model shows some deficiencies, in particular by lacking values for some tectonic blocks, which have been clearly identified in recent years as separate units. In addition, it only provides estimates of the angular velocities for 14 large

plates when more recent research shows a significant larger number of independent tectonic units. This is the case of PB2002 (Bird 2003), which is a compilation of a total of 52 tectonic plates.

However, even this figure is not final. With the densification of observation systems, more plates have been identified. A good example is the African continent, which was considered to be a single unit in NUVEL-1A and in PB2002 was divided into two blocks (Nubia and Somalia) as depicted in Fig. 5.2. In a more recent publication, Stamps et al. (2008) have identified (and quantified) another separate unit, namely the Victoria Block. The availability of space techniques allowed an assessment of the actual plate movements and brought in new information that was highly relevant for the clarification of some of these aspects.

The global models of plate motions, based on geophysical and geological data averaged over the past few million years, are a useful reference for comparison with motions estimated from space-geodetic measurements, averaged over the past few decades (not more than one decade for many plates). However, a timescale issue has to be considered when comparing current day plate boundary configuration from GPS measurements with that deduced from plate tectonic models that integrate the past few million years. For many plates it has been observed that a significant distinctive motion between geological/geophysical and geodetic predicted angular velocities exists (Norabuena et al. 1998; Sella et al. 2002). This probably reflects a

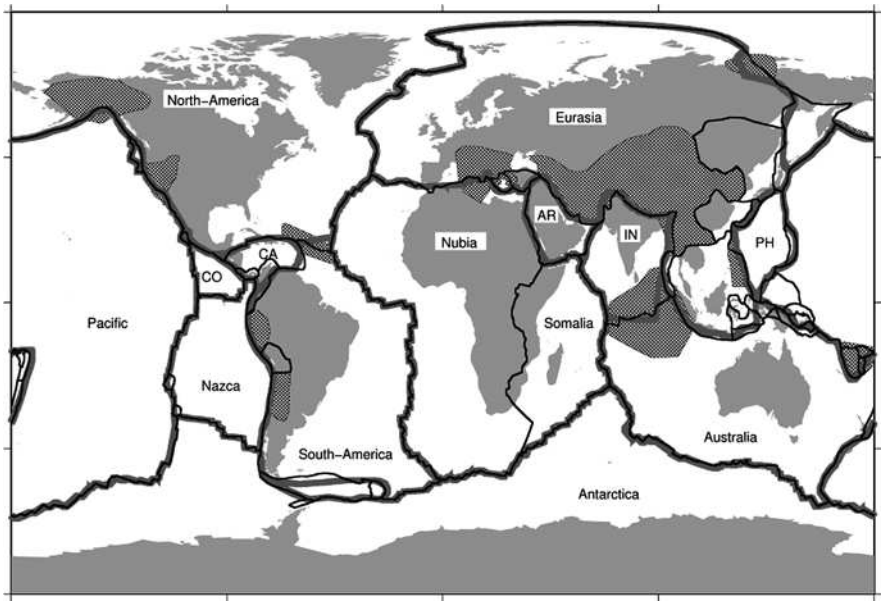


Fig. 5.2 The boundaries of the tectonic plates of the NUVEL-1A model (DeMets et al. 1994) are shown by the *thick gray lines*. The plate boundaries of the PB2002 model (Bird 2003) are represented by the *thin black lines*. Wide deformation areas (orogens) are shown with the *squared pattern*

change in the angular velocities due to processes not clearly understood at present. Plate tectonic models can also be biased by local effects (inter- and intra-plate movements), and reveal inconsistencies, especially in boundary zones. GPS observations give motions over a short time interval, a few years, showing aspects of present-day plate tectonics with consequences in the definition of boundaries and present-day behaviour. In other words, if the secular motions of tectonic plates can change over time, then the geophysical/geological models will provide the average motion over a few million years, and consequently they will not depict exactly the present-day velocities, whereas geodetic-based models provide estimations of the secular motions based on only a few years. Nevertheless, geophysical/geological models like NUVEL-1, and successors, still provide useful information to study present-day kinematics in areas, such as oceans, where geodetic measurements are not available. On the other hand, GPS solutions in deformation zones within the continents are providing more reliable information about the kinematic constraints than that provided by geological/geophysical plate motion models. This information is also contributing to revealing to what extent inter- and intra-plate deformations occur in a continuous or episodic manner.

The coverage provided by space geodesy, with networks implemented worldwide, gave a global perspective that allowed the establishment of new geodetic-based plate models such as Larson et al. (1997), REVEL (Sella et al. 2002), DEOSVEL (Fernandes et al. 2003) and ITRF2005 (Altamimi et al. 2007). Since the mid-1990s these models are being continuously improved, both in the number of described plates and in the number and length of the time-series used. Table 5.1 illustrates the previous discussion by showing the differences in the estimated angular velocities and position of the rotation vector on the Earth's surface (also known as the Euler pole) provided by the different models.

The values for three major plates, North America, Eurasia and Nubia (western part of Africa), are presented in Table 5.1. The boundaries of the North American plate are well identified, whereas the limits of the Eurasian and Nubian plates changed significantly in the last years with the identification of other blocks.

The comparison of the solutions clearly shows that the estimated angular velocities for most plates are currently reaching robust values. With the exception of AM1-2, the predictions provided by all models for these plates are similar. Notice that no uncertainties are provided with the values displayed in Table 5.1. The reason is that the associated uncertainties provided by most models do not reflect the real uncertainty in the computed values. The published values are usually too optimistic since they were computed using the formal errors that are directly obtained from the least-squares approach used to compute the time-series. When least-squares are used, it is normally assumed that only white noise processes affect the coordinate positions, which is not true since they are also affected by a multitude of systematic errors (monument, orbits, etc.). This is discussed in more depth in Sect. 5.5. As a consequence, the formal uncertainty only depends on the number of used position solutions and their associated formal error. It does not transmit how good the data fit the model (in this case, how close the daily solutions are to the estimated secular trend).

Table 5.1 Solutions provided by different models: AM1-2 (Minster and Jordan 1978), NUVEL-1A (DeMets et al. 1994), Larson et al. (1997), REVEL (Sella et al. 2002), DEOSVEL (Fernandes et al. 2003), ITRF2005 (Altamimi et al. 2007) for three major plates North America, Eurasia and Nubia (Africa = Nubia + Somalia in AM1-2 and NUVEL1-A). Latitude is denoted by φ, longitude by λ and the angular velocity by ω

Solution	North America	Eurasia	Nubia (W. Africa)
AM1-2	$\varphi = -58.31°$	$\varphi = 0.70°$	$\varphi = 18.76°$
	$\lambda = -40.67°$	$\lambda = -23.19°$	$\lambda = -21.76°$
	$\omega = 0.247°$/Myr	$\omega = 0.038°$/Myr	$\omega = 0.139°$/Myr
NUVEL-1A	$\varphi = -2.429°$	$\varphi = 50.655°$	$\varphi = 50.656°$
	$\lambda = -86.035°$	$\lambda = 112.562°$	$\lambda = -74.081°$
	$\omega = 0.2064°$/Myr	$\omega = 0.2336°$/Myr	$\omega = 0.2906°$/Myr
LARSON	$\varphi = -0.4°$	$\varphi = 56.3°$	$\varphi = 50.0°$
	$\lambda = -84.5°$	$\lambda = 102.8°$	$\lambda = -86.8°$
	$\omega = 0.22°$/Myr	$\omega = 0.26°$/Myr	$\omega = 0.26°$/Myr
REVEL	$\varphi = -2.39°$	$\varphi = 58.27°$	$\varphi = 52.25°$
	$\lambda = -79.08°$	$\lambda = 102.21°$	$\lambda = -80.18°$
	$\omega = 0.199°$/Myr	$\omega = 0.257°$/Myr	$\omega = 0.253°$/Myr
DEOSVEL	$\varphi = -4.574°$	$\varphi = 54.614°$	$\varphi = 50.861°$
	$\lambda = -83.150°$	$\lambda = -103.876°$	$\lambda = -81.475°$
	$\omega = 0.1945°$/Myr	$\omega = 0.2485°$/Myr	$\omega = 0.2610°$/Myr
ITRF2005	$\varphi = -4.291°$	$\varphi = 56.330°$	$\varphi = 49.955°$
	$\lambda = -87.385°$	$\lambda = -95.979°$	$\lambda = -82.501°$
	$\omega = 0.192°$/Myr	$\omega = 0.261°$/Myr	$\omega = 0.269°$/Myr

Bastos et al. (2005) showed that, when enough GPS solutions exist (as is the case for the plates shown in Table 5.1), the implication of choosing different error models in the computation of the velocity solutions used to derive the angular velocity model is not significant. In particular they analysed the implications of the chosen noise model on the estimation of the angular velocities for the Nubia and Somalia tectonic plates using a network of continuously operating GPS stations.

They used three different error models: the usual formal uncertainties (directly obtained from the least-squares adjustment); the WRMS of the residuals divided by the observation span in years (Fernandes et al. 2004), which is an empirical formulation; and a power law + white noise model (Williams et al. 2004), which takes into account the existing noise signals in the trend.

The differences between the estimated angular velocities for Nubia (14 sites) using the three approaches were smaller than 0.7º and 0.003º/Myr. These differences were significantly larger for the Somalian plate, where only four sites were available, reaching 3.8º and 0.010º/Myr. However, the implications on the predictions given by the tectonic models are not very significant, as can be observed in Fig. 5.3 which shows the relative motion of the stable part of the Somalian plate with respect to Nubia (Bastos et al. 2005).

It is worth mentioning that including a seasonal signal during the estimation of the GPS velocity vectors can change the Euler pole by up to 0.5º and 0.002º/Myr.

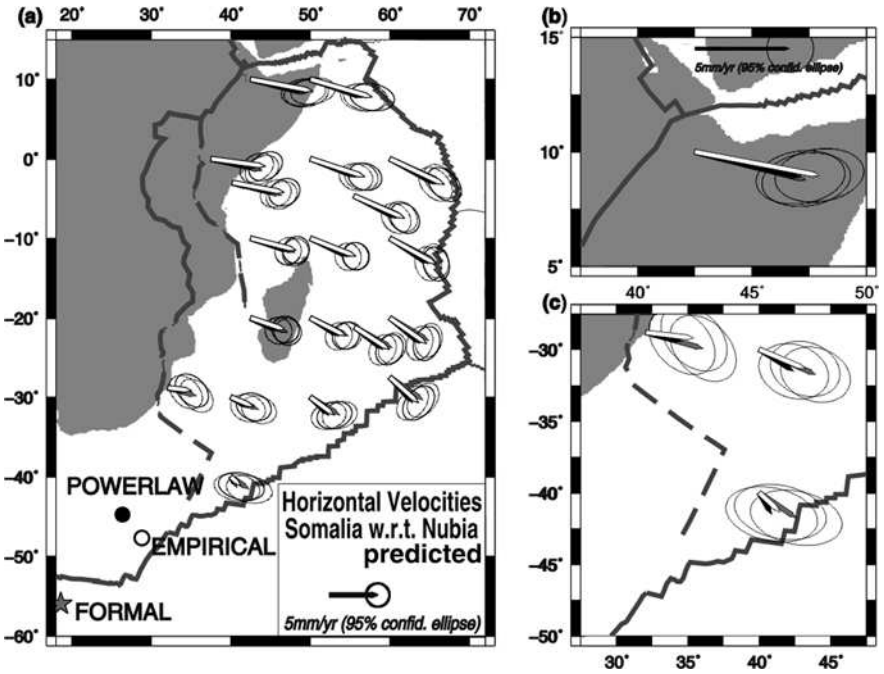


Fig. 5.3 Comparison between the relative angular velocity of Somalia with respect to Nubia by using different noise models for the used horizontal velocities (Bastos et al. 2005)

The definition of a plate model is also relevant for the realisation of a Reference Frame as it is used as its kinematic reference. The model should fulfil the condition of no-net-rotation which states that the integral over the Earth's surface of all plate motions should be zero. To perform this integral it is necessary to know the plate boundaries which are inferred from geological/geophysical information. The current practice is to align the geodetic plate models to a geological/geophysical plate model to ensure that the no-net-rotation condition is satisfied (Kreemer et al. 2006).

5.3 Mapping Issues

To align daily (or combined) GPS solutions with any realisation of the International Referential Reference System (currently, ITRF2005 – Altamimi et al. 2007), the usual first requirement is to process the network of interest together with a sub-set of stations with well-known position and velocity estimates in the desired reference frame. These reference stations are then used to compute the transformation parameters to project the daily (or combined) "unknown" reference frame onto our reference frame.

The selection of the reference stations is a fundamental step of the mapping process. Although ITRF2005 represents a large improvement in comparison with the previous ITRF versions, the amount and distribution of GPS stations (typically IGS stations) with well-determined ITRF2005 position/velocity representation is still not optimal. The problem is noteworthy for many regions since, although the existence of a significant improvement was seen in recent years, more than half of the reliable IGS stations are still located in Europe and North America.

We demonstrate here the implications of using different sets of reference stations to align the daily solutions into the reference frame of interest. Different sub-sets of regional stations have been used in Europe to compute the daily transformation parameters since January 1996. Figure 5.4 shows the distribution of the selected stations. First, a global distribution of stations was considered with a spatial distribution as uniform as possible. Second, for the regional reference network, the same sub-set of 13 stations chosen by EUREF (EPNCB 2008) to map the EPN weekly solutions into ITRF2005 has been selected. EPN is a dedicated network created to define a unified European reference frame (ETRS89) (Bruyninx 2000). It currently counts approximately 200 stations concentrated in Europe and

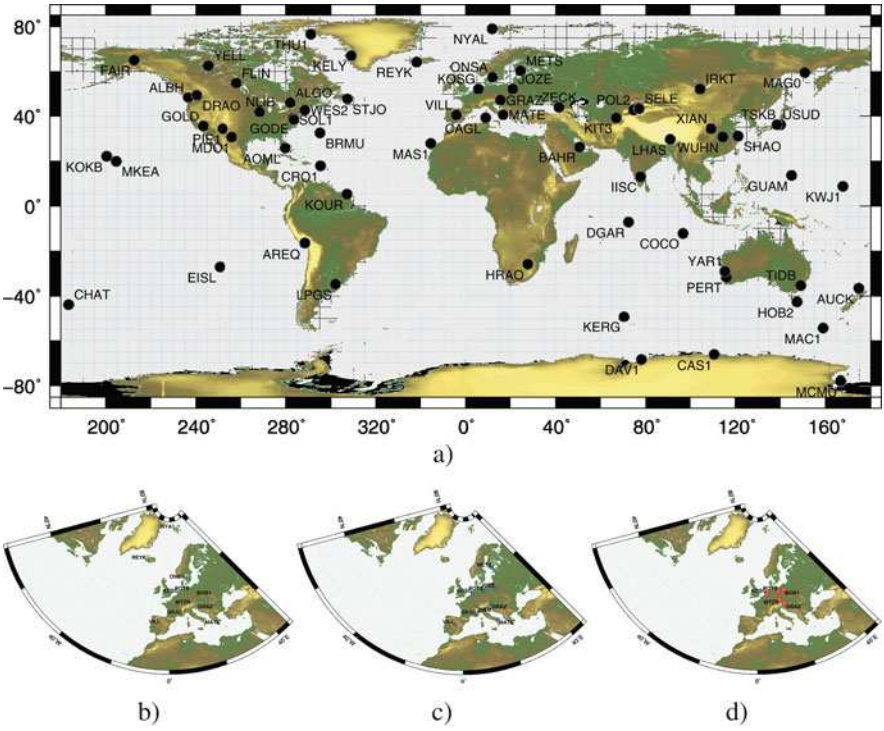


Fig. 5.4 Four different configurations of reference networks (RN) used for these tests: global RN (**a**); RN used to compute the EPN weekly solutions (**b**); RN used by the EUREF Special Project "Time-Series" (**c**); a sub-regional network located in the Central Europe (**d**)

surrounding regions. Third, since 1996, different sets of reference stations have been used to tie the EPN combined solution to the successive ITRS realisations. The last set was selected since it is assumed that this set has been fully tested and provides the most accurate link of the EPN stations into ITRF2005 on a European scale. We will call this the "Regional Time-Series" set since they were used by EPN in a special project to compute time-series. Fourth, a sub-set of five reference stations localised on Central Europe was selected to compute another set of transformation parameters, considered here as a sub-regional reference network.

These different reference networks were used to compute time-series for KOSG (Kootwijk, the Netherlands). This station has been selected since an unexpected event occurred at this station in 2003 (the mast was hit by a car causing a jump on the time-series). In fact, the analysis of the derived time-series (see Fig. 5.5) illustrates some important issues:

- The jump due to the event is clearly visible in the Global/EPN weekly/Regional Time-Series.
- However in the sub-regional network, that jump is masked by the few number of stations used in the mapping. Furthermore, the time-series for KOSG, particularly for the sub-regional, are dictated by the fact that this station is a reference station among five (also more visible in the EPN and Regional Time-series than in the global mapping).
- Periodical signals in the time-series are better visible (and cleaner) on the global solution than in the other solutions.

The consequences of using different types of reference networks are also evident in Fig. 5.6. It depicts the solutions for Maspalomas (MAS1), located on Canary Islands, i.e. outside of the limits defined by the EPN Weekly, Time-Series project, and sub-regional reference network stations. Consequently, the computed transformation parameters were extrapolated for this station when these networks were used. The degradation in the estimated signal is clearly observed. In fact, the amplitude of the noise prevents to compute any reliable solution for the three components of the position when the sub-regional network is used. The EPN Weekly and Time-Series project solutions are also clearly noisier than the global solution.

The antenna accident at KOSG also reveals the consequences of the use of a small number of reference stations. No correction has been added intentionally. Two facts are evident from the comparison of the time-series between the global and the sub-regional solutions: as already mentioned, the signal is well identified on the global solution but it is almost not detected on the sub-regional solution, independently of the fact that KOSG is used in both solutions as reference station.

Inversely, no consequences are observed in the global time-series of MAS1, whereas all weekly sub-regional solutions for MAS1 are considered outliers after that date.

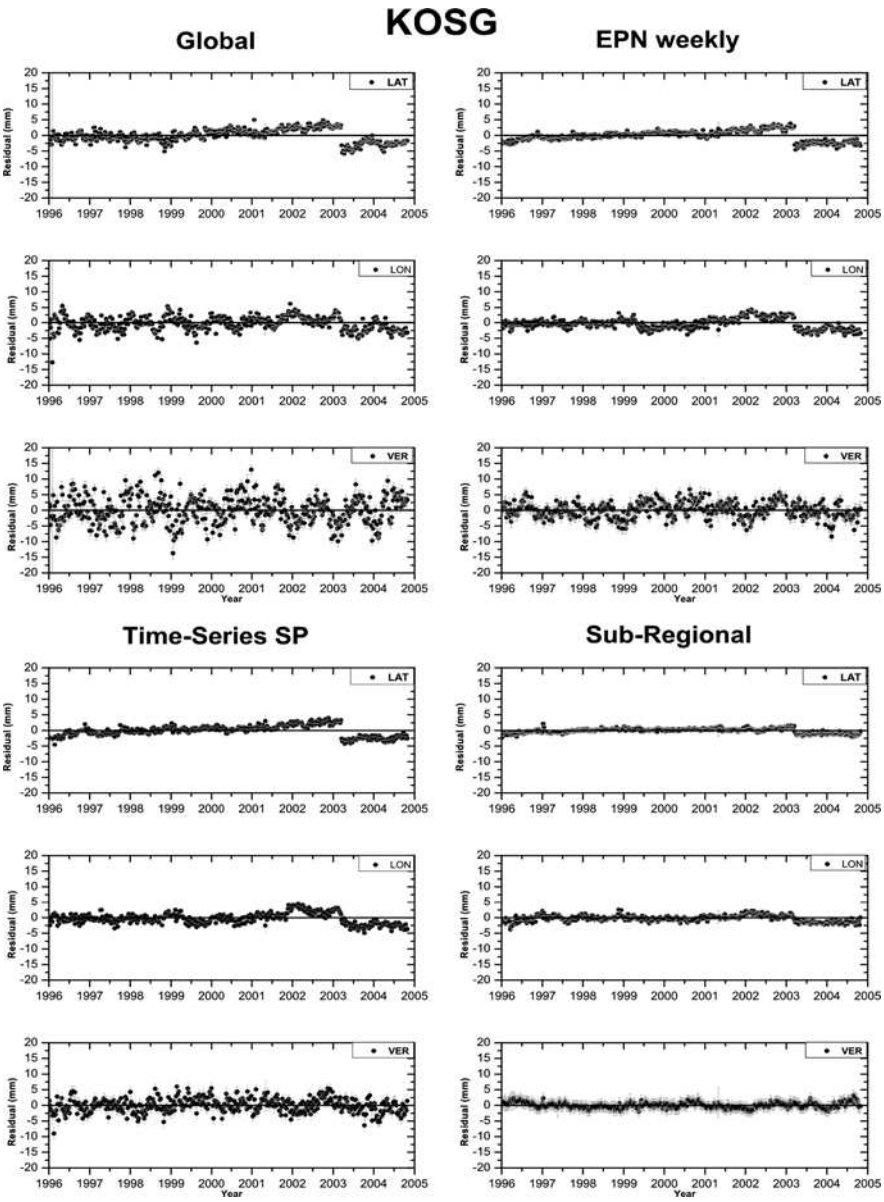


Fig. 5.5 Time-series of KOSG using different mapping reference networks

The conclusion is that the number of stations influences significantly the estimation of the transformation parameters; when a small number of stations are used, an error on one station position affects the entire solution, whereas those effects are mitigated when a large number of stations are used (more than 60 stations were used on the global solution).

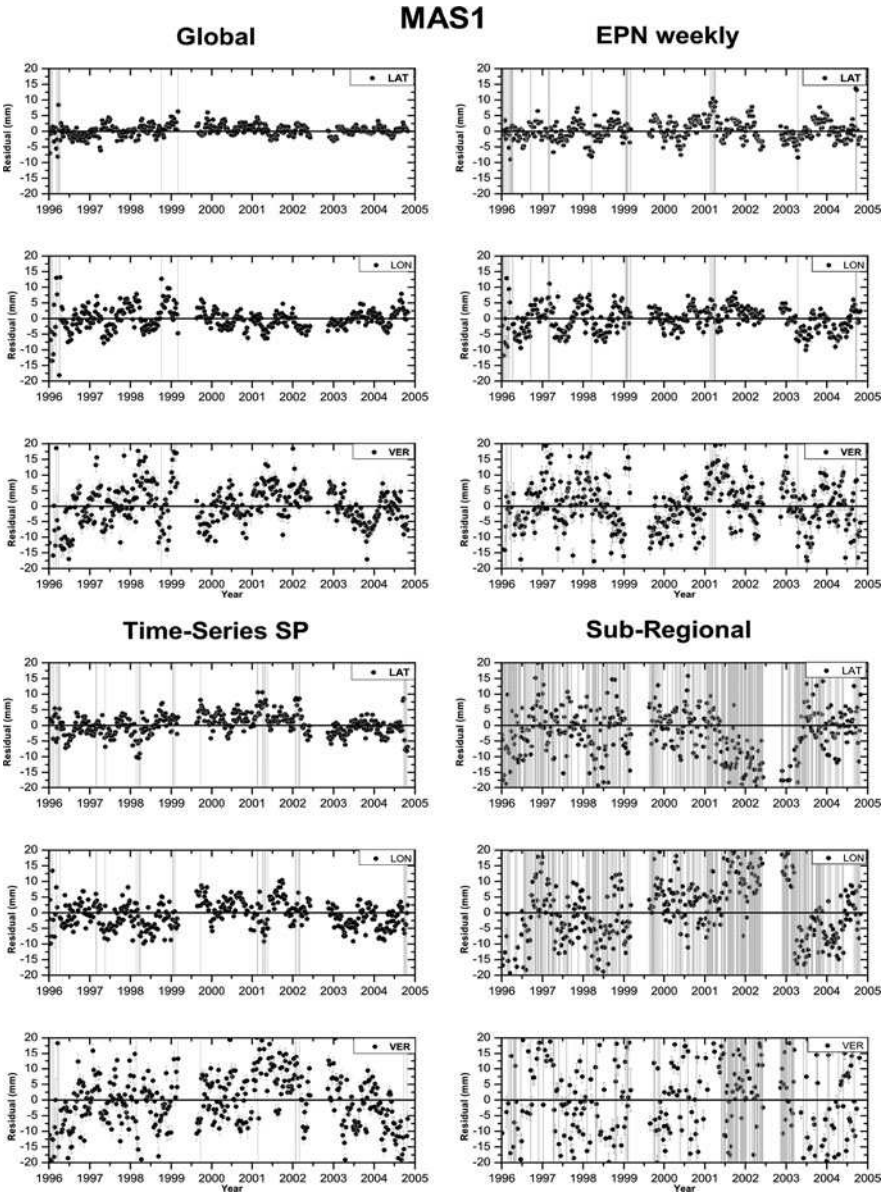


Fig. 5.6 Time-series of MAS1 using different mapping reference networks

5.4 Geophysical Corrections for the GPS-Derived Station Positions

In the previous section we described how to map GPS solutions for the position of a station over time within the International Terrestrial Reference Frame (ITRF). Each particular realisation of the ITRF, such as ITRF2000 (Altamimi et al. 2002) or more recently ITRF2005 (Altamimi et al. 2007), consists of a global set of geodetic stations for which one defines a position $\overline{X}_0$ at time t_0 and a velocity $\overline{V}_0$. These values represent our best fit of the tectonic plate motions to our observations. Consider now the following equation:

$$\overline{X}(t) = \overline{X}_0 + \overline{V}_0 \cdot (t - t_0) + \sum_{i=1}^{N} \Delta\overline{X}_i(t). \tag{1}$$

Here $\overline{X}(t)$ is the instantaneous position of the geodetic reference station that is observed with GPS, given in ITRF coordinates, at time t. The last summation contains all the geophysical corrections that should be added to the plate motion to obtain the real instantaneous position and these are the subjects of this section. The conventions of the International Earth Rotation Service (IERS) prescribe which geophysical corrections should be applied during the GPS analysis to ensure homogeneity between all published GPS results (McCarthy and Petit 2004). If these geophysical corrections are subtracted from the estimated instantaneous position, one obtains a mean station position that should correspond better to the tectonic plate motion one wants to investigate. The current custom is to analyse batches of 1 day of GPS data which means that one obtains a time-series of mean daily positions.

The largest geophysical correction is the tidal deformation of the solid Earth, also called the body tide, caused by the gravitational attraction of the Moon and Sun. It produces periodic variations of tens of centimetres in the position in a time span of several hours. The current model of the body tide assumes that the deformation of the Earth is linear with the tidal forces and that it is for around 99% instantaneous. For the most dominant period of 12.42 h, called harmonic M_2, the body tide lags with $0.204° \pm 0.047°$ (Ray et al. 2001) behind the tidal forcing. The elasticity of the Earth is also dependent on the period of the tidal forcing, especially near the resonance period of the free core nutation, which is close to one sidereal day. Despite these complexities, the body tide can be modelled to better than 1 mm accuracy (McCarthy and Petit 2004; Watson et al. 2006).

The second largest geophysical correction is due to the loading of the ocean tides on the ocean floor. This deformation continues land inwards and can reach centimetre level. For the dominant tidal harmonic M_2 the accuracy of this correction is mostly smaller than 0.4 mm for inland sites but can reach up to 3 mm at some coastal sites.

The Earth not only deforms due to the tidal forces of the Sun and Moon, but also changes its orientation in space. This is called nutation, and IERS prescribes a

very accurate model to correct for this (McCarthy and Petit 2004). The ocean tides also change the orientation of the solid Earth because the total angular moment of the whole Earth system must remain constant over time. Therefore, a change in the angular momentum of the oceans, due to tidal currents and change in tidal elevation, must be compensated by a change in the rotation of the solid Earth. The effect can be as large as 10 mm and the model uncertainty is about 0.2 mm, mostly in the horizontal position.

Since tides have their largest influence at the semi-diurnal and diurnal periods, their influence on the secular motion of the station is small because most of this signal averages out over a period of 1 day. However, studies by Stewart et al. (2005) and Penna et al. (2007) show that one still needs to be cautious and that one must apply accurate tidal corrections since any error in these corrections can cause significant spurious periodic signals at longer, such as annual, timescales.

The orientation of the Earth in space is also affected by changes in angular momentum in the atmosphere and due to non-tidal currents (Gross et al. 2003). This is called polar motion and at the moment no model exists to predict it; only observed values are available that must be inserted into the GPS analysis. Polar motion has its largest effect at the yearly period and the Chandler period which is around 14 months. The maximum radial displacement due to polar motion is approximately 25 mm, and the maximum horizontal displacement is about 7 mm.

The atmospheric pressure also loads and deforms the Earth surface in the same way as ocean tide loading and causes deformations that can reach several millimetres. Atmospheric loading has its largest effect at the annual period which means that this effect does not average out as quickly as the tidal corrections. The atmospheric loading values are provided by the Special Bureau for Loading (http://www.sbl.statkart.no/). The uncertainty of the atmospheric corrections is mostly due to the errors in the surface pressure data and is estimated to have an RMS of 0.75 mm and a maximum value of 3 mm. An extra complication is the dynamic response of the oceans due to variations in the atmospheric pressure. One mostly assumes that the response is instantaneous and that the sea level is inversely proportional to the pressure changes above it. This is called the inverted barometer assumption but it is only valid for periods longer than 10–30 days.

In addition, ground water level and thermal expansion can also induce variations in the stations position, mostly with a period of 1 year (Zerbini et al. 2001). However, no conventions exist to deal with these phenomena. Residual annual signals are mostly estimated when analysing the GPS position time-series as explained in Sect. 5.5.

Since the last ice age, some 20,000 years ago, the Earth's crust is still slowly recovering in some places from the deformation caused by the kilometres thick ice layers that were covering areas such as Scandinavia and Canada. The vertical uplift of this post-glacial rebound can be as large as 1 cm/year (Scherneck et al. 2003). There are also horizontal deformations of several millimetres per year associated with this phenomenon because the uplift causes a horizontal flow of mass. There exist models for this post-glacial rebound effect but these are not part of the standard corrections that are advised by IERS.

Finally, corrections are available for the effect caused by the movement of the centre of mass of the solid Earth with respect to inertial space. According to IERS, the ITRF is defined to have its origin in the centre of mass of the solid Earth. Due to the mass distribution in the ocean due to tides, and to the non-tidal currents and mass distributions in the atmosphere due to weather, the solid Earth moves several millimetres with respect to an inertial reference frame to keep the centre of mass of the whole system constant. The satellites that support space-geodetic techniques, such as SLR, DORIS and GPS, orbit around the centre of mass of the whole Earth which includes the atmosphere and the oceans. For this reason, IERS prescribes to use atmospheric and ocean tide loading corrections that have the associated geocentre motion added to them. Although the geocentre motion of several millimetres, mostly at the yearly period, has been clearly observed with SLR, DORIS and GPS (Chavet et al. 2003; Feissel-Vernier et al. 2006), the differences between the space-geodetic techniques are significant and further research is needed to resolve these discrepancies.

5.5 Time-Series Analysis

After obtaining a time-series of mean daily positions derived from GPS observations, one can estimate the station motion from these data. This motion is thus assumed to be equal to the plate motion on which the station is located. An example of such a time-series is given in the top panel of Fig. 5.7 which contains the variation in the position towards the north at KOSG given in IRTF2000. In the top figure

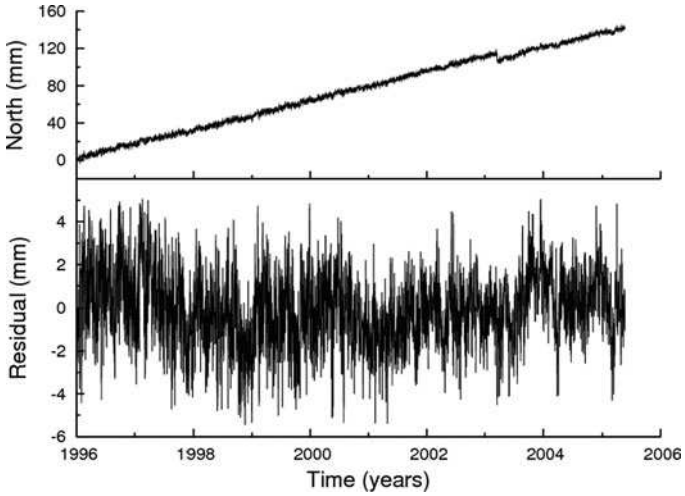


Fig. 5.7 The variation of the GPS positions over time at KOSG, north component. The *top panel* shows the original time-series while the *bottom panel* shows the time-series after subtracting a linear trend and a correction for the jump in the data

one can see clearly a linear motion over the years. In addition, in 2003 there was, as mentioned before, an accident which caused the antenna to be displaced by several millimetres.

The KOSG station is a core station of ITRF2000. Therefore, by definition the north velocity of this station should be exactly equal to 15.3 mm/year. As was described in Sect. 5.3, due to measurement errors and errors between the real and defined ITRF velocities, this fit is not perfect and one obtains time-series like that presented in Fig. 5.7 with a trend value of 15.76 ± 0.14 mm/year.

The standard method for estimating this linear motion from these data is ordinary least-squares. However, ordinary least-squares only gives optimal results when all observations are independent and normally distributed. One of the first papers that emphasised that the assumption of independent GPS observations is invalid was by Johnson and Agnew (1995). To be precise, after subtracting the estimated linear motion from the estimated GPS positions, one obtains a time-series of residuals that is assumed to represent the noise in the observations. If the observations were independent, the correlations between the residuals should be zero which is not the case. The work of Johnson and Agnew was the result of earlier studies by Agnew (1992) who showed earlier that temporal correlations also exist in tide gauge and laser strain data. Recently, temporal correlations have also been identified/demonstrated in time-series of absolute gravity measurements (van Camp et al. 2005). Therefore, it is more unusual to find a geodetic time-series that does not show temporal correlations than to find one which does. In addition, the temporal correlations have been shown in all cases to decrease only very slowly with increasing time span between each pair of observations and are for that reason called long-range correlations (Beran 1994). Since it is assumed that the noise at time t depends on a weighted sum of all previous noise values, these long-range correlations can produce a noise that is not stationary; actually it gets larger over time. The most famous example is random walk behaviour which has actually been observed in two colour electronic distance measurements (Langbein 2004). In GPS observations, the long-range correlations are less strong but still cause the noise to grow on average over time instead of being a stationary signal.

The non-stationary noise can easily be confused with the tectonic motion that one wants to investigate. This is the reason why the real uncertainty in the estimated tectonic motion can be much larger than that predicted by using ordinary least-squares which ignores this effect. For many years, most researchers using ordinary least-squares were well aware that their error bars were too small. To overcome the associated consequences they applied a scaling factor that used the variance of the GPS residuals as observation error instead of those provided by the GPS analysis software. However, Mao et al. (1999) showed that this can still cause an underestimation of the real error in the trend by a factor of 5–11 because one still assumes that the observations are independent. Bos et al. (2007) showed that the underestimation is even slightly larger, 6–13 times. A realistic error bar on the estimated trend value is extremely important when one tries to determine whether two sets of estimated plate motions are significantly different.

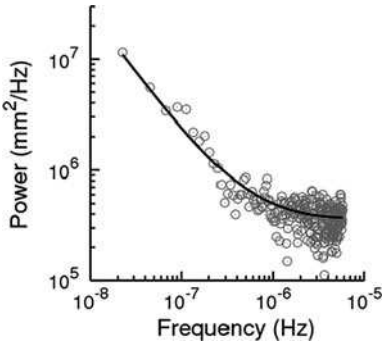


Fig. 5.8 The power spectrum of the GPS north residuals at KOSG (*grey circles*). The *solid line* represents the power law plus white noise model fitted to the power spectrum

One method to investigate the temporal correlations is by looking at the power spectrum of the residuals. Figure 5.8 shows the power spectrum computed using the GPS residuals of the north component at the KOSG station.

A time-series of independent random variables has the same power at all frequencies. This is also called white noise. One can see that the high frequencies in Fig. 5.8 are well described by a constant power. On the other hand, the power of the lower frequencies in Fig. 5.8 seems to increase exponentially with decreasing frequency. This is called power-law behaviour and is the result of the long-range correlations in the GPS observations. In mathematical notation this behaviour of the power spectrum can be expressed as

$$P(f) = P_0 \left(\frac{1}{f}\right)^{\alpha}, \tag{2}$$

where P_0 is a constant, f is the frequency and α is called the spectral index (Kasdin 1995).

The first authors to present this type of evaluation of the power spectra for GPS observations were Zhang et al. (1997) and Mao et al. (1999). Mao et al. also showed that maximum likelihood estimation (MLE) provides the most accurate estimate of the trend and the values of the noise parameters. In their case these noise parameters were the variance of the white noise and the variance of the power-law noise. The value of their spectral index of the power-law noise was fixed to one which is also called flicker noise. Williams (2003a) generalised the noise model by also estimating the spectral index α value in the MLE process. However, experience has shown that the spectral index of the power-law noise component in all GPS time-series is always close to one and that therefore fixing it a priori to this value is generally an acceptable practice (Williams et al. 2004). No explanation for the fact that the spectral index is always around one has yet been given.

The MLE provides the values of the trend motion and the noise parameters that are the most likely to have occurred for the given data set. Since the problem is not linear, due to the noise parameters, one must use a numerical maximisation algorithm to determine the best likelihood value (Williams 2003a; Williams et al. 2004; Langbein 2004; Bos et al. 2007).

The MLE described above is implemented in the CATS software (Williams 2008) that is becoming a popular choice for analysing GPS time-series, being used for example by EUREF in the analysis of their permanent network.

Besides temporal correlations, there exist correlations between the time-series of the North, East and up component. However, these correlations are mostly neglected because the effect of the temporal correlations on the accuracy of the estimated trend values is much larger.

In addition, there exist spatial correlations between the time-series of stations that are closely located to one another (Wdowinski et al. 1997). As an example we present in Fig. 5.9 the spatial correlation between the residual time-series, north component, of 24 permanent GPS stations located on the Iberian Peninsula.

One can see that the correlation is significantly different from zero and that the correlation decreases with increasing distance between the stations. The exact cause of this correlation is still unknown but it could be due to errors in the orbit which propagate into similar station position errors.

Figure 5.10 shows that taking those correlations into account has an impact on the computation of the velocity field for the region.

It has already been mentioned that the spectral index of the power-law noise has always been observed to be close to one. Williams et al. (2004) investigated whether the variances of the power law and white noise depend on the type of geodetic monuments. This is an important topic since reducing noise levels in geodetic data is critical for the interpretation and modelling of geophysical interesting signals. As was to be expected, the highest variance was observed for GPS antenna's installed on oil platforms. The second worst type of monument was the common concrete pillar. However, this last result has been contradicted by Beavan (2005). He concluded that differences between the monument types are not a dominant influence on the noise, implying that the relatively cheap concrete pillar monuments are still a valid choice.

Unfortunately, most GPS time-series contain outliers which, by using common sense, should be removed from the data set in order to avoid getting nonsense results. A simple but robust method for outlier removal is to compute the median and the 25 and 75% interquartile. Anything that is smaller than three times the value of the

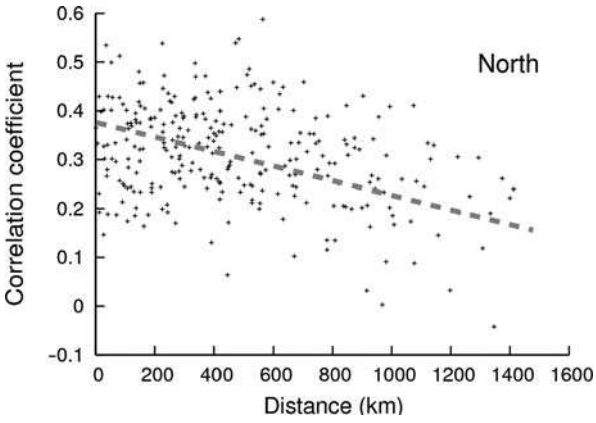


Fig. 5.9 The spatial correlation in the north component between all pairs of residual time-series of 24 permanent GPS stations located in the Iberian Peninsula as function of the distance between the stations

Fig. 5.10 Velocity field (w.r.t. Nubia) before (*grey*) and after filter (*white*) (Bastos et al. 2006)

median minus the 25% interquartile and anything that is larger than three times the 75% interquartile minus the median is considered to be an outlier.

In addition, one has to look for offsets in the data and tell the analysis software that, on that particular epoch, it must also estimate an offset. Currently, the time at which offsets occur is mostly determined by visual inspection of the time-series. Williams (2003b) has shown that it is virtually impossible to detect all offsets in the time-series because some could be smaller than the noise. A result could be that over very long time spans the noise in the GPS time-series will be dominated by random walk noise, caused by the accumulation of offsets, instead of flicker noise. However, even at stations with more than 10 years of GPS data, no random walk noise has yet been observed which assures us that the stability of most GPS monuments is quite robust over the years.

At the plate boundaries one has to check whether the tectonic motion is the same before and after an offset. When the source of the accident is known to be a displacement of the GPS antenna, as was the case at KOSG, the tectonic motion will clearly be unaffected. However, in other cases, such as station LDES in California, USA, which is located at a plate boundary, one has observed a quite distinct change in the velocity after an earthquake. Figure 5.11 shows the time-series of the north component for the continuous GPS station LDES. The Hector Mine earthquake in 1999 caused an 18-cm offset in the north component which has already been corrected for. Next, one can see some period after the earthquake where the position seems to experience an exponential relaxation before assuming a new linear tectonic motion (Pollitz 2003).

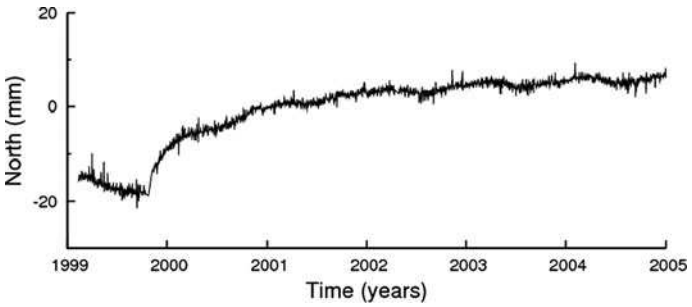


Fig. 5.11 The GPS position, north component at the LDES station (SOPAC Refined Model GPS site Position Time Series, http://sopac.ucsd.edu)

So far we have only discussed continuous GPS observations. However, in some areas only GPS campaign data are available with intervals of around 1 year or more. Time-series of several years of GPS campaign data can clearly show the tectonic motion as was described in Sect. 5.1. On the other hand, the large gaps in time between each campaign severely limit the application of the MLE method described in this section to determine the noise properties because there is mostly too little data available. This is still an area of current research but one could probably assume a conservative a priori power law plus white model which is observed at a nearby permanent station and use this model in the trend estimation from the GPS campaign data.

5.6 GPS and Geodynamics – An Example

We present in this section an example of the use of GPS for studying the Earth's deformation due to geodynamic processes in the Azores Triple Junction Region. Such an example does not intend to give an exhaustive representation of all the research done in this field since it is already more than 20 years that GPS has become the main tool to directly observe present-day tectonic motions.

Most of the initial GPS-based geodynamic projects were developed at a regional level. Among others, we can refer the works from Dong and Bock (1989) and Larsen and Reilinger (1992) in the California region; Straub and Kahle (1995), Noomen et al. (1996), Reilinger et al. (1997) and Kahle et al. (1995) in the Mediterranean area; and Feigl et al. (1993), Tsuji et al. (1995) and Miyazaki et al. (1996) in Japan. In parallel, projects with a broader geographical (intercontinental) scope also started to be developed. Examples are the CASA UNO (Kellogg and Dixon 1990) in Central and South America and the TANGO (Bastos et al. 1991) in the Azores-Gibraltar region, which were among the first GPS networks established to support geodynamic studies.

The TANGO project, established in 1988 with GPS stations in the European and North American continent, included a network on the Azores Archipelago, shown

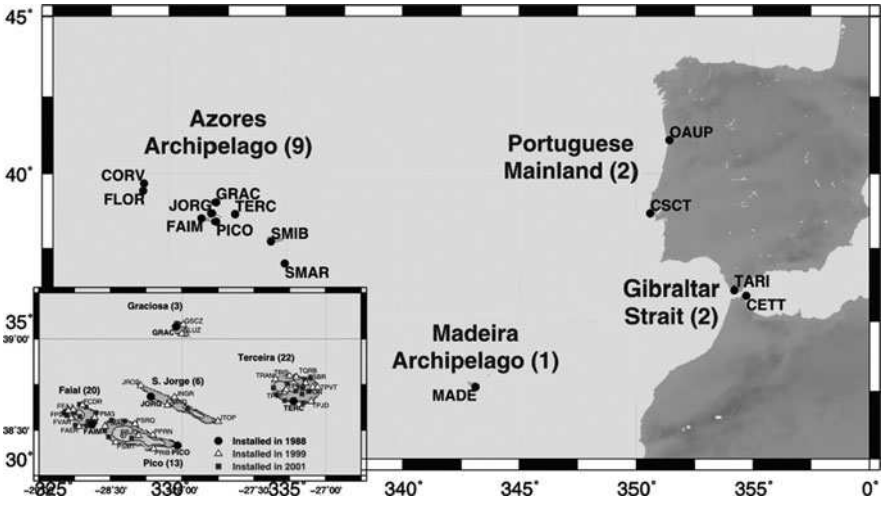


Fig. 5.12 TANGO network installed in 1988. The *inset* shows the stations installed on the five islands of the Central Group of Azores Archipelago in the framework of other collaborating projects since 1999

in Fig. 5.12, located near the mid-Atlantic ridge in the area of the so-called Azores Triple Junction (ATJ). The object of this project was to contribute to the description of the complex tectonic features of this region where three large tectonic plates meet: the North American, Eurasian and Nubian plates. Several authors have tentatively drawn a geodynamical model for the ATJ aiming to resolve the configuration of the boundary between the Eurasian and Nubian plates.

The spatial distribution of the deformation associated with the Eurasia–Nubia plate boundary is as yet poorly understood. The fine scale geometry of this active plate boundary area is still unclear, and the information derived from the GPS observations is giving a unique and decisive contribution to understand its broad structure.

Presently, a record of 20 years of periodic GPS campaign data in the Azores area exists. Complementary projects have meanwhile been developed (Navarro et al. 2003; Fernandes et al. 2004) providing a much better spatial coverage, in particular in the Central Group, with the installation of dense campaign networks per island between 1999 and 2001. The TANGO network is now quickly evolving towards a permanent approach.

Results from the initial TANGO network for the period 1988–1997 were presented by Bastos et al. (1998), showing accuracy at the few centimetre level. Consequently, a period of about 10 years with periodic campaigns was thought to be sufficient to draw definitive conclusions about the intensity and direction of the plate motion in the different islands.

This is clearly depicted in Fig. 5.13 extracted from Fernandes et al. (2006), which shows the horizontal velocities at the Terceira, Pico and S. Jorge islands. It is possible to observe large differences between the velocities of the stations located on

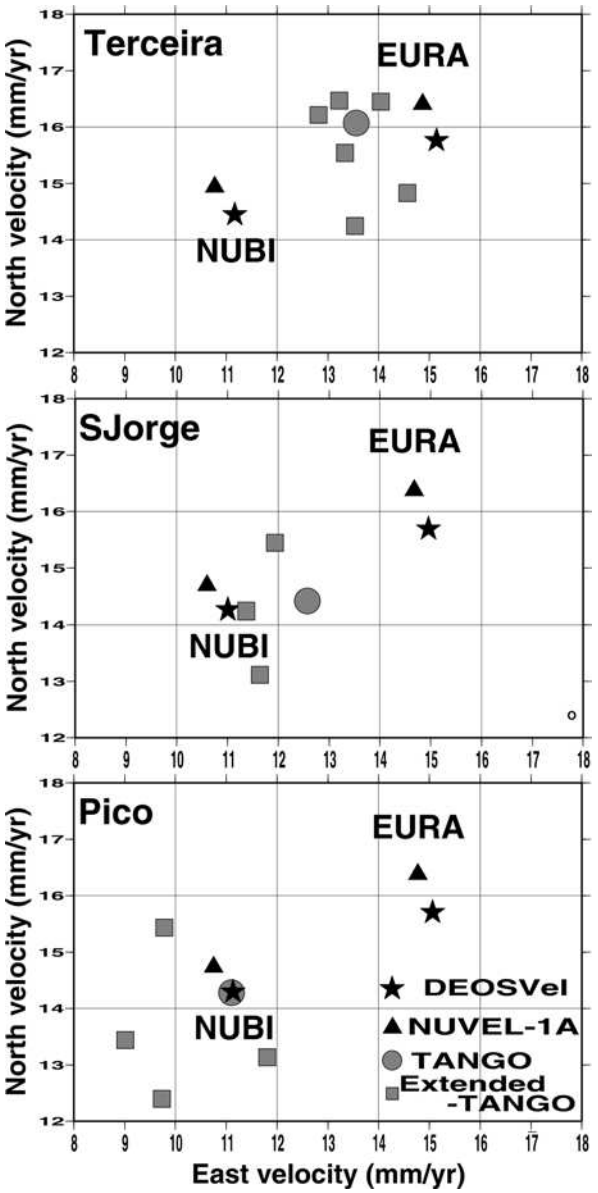


Fig. 5.13 Horizontal motions for the three islands in Central group. Estimated motions for initial sites (TANGO *circles*) and recent sites (Extended-TANGO, *squares*). Also drawn are the predicted motions according to the DEOS2k (*stars*) plate model and the NUVEL-1A (*triangles*) plate model

the same island. The distribution of these differences is random, without a correlation between observed motions and known tectonic features. However, the averaged absolute velocity of all these velocities on each island is close to the solution given for the TANGO site at each island, respectively.

The conclusion is that the TANGO sites already provide enough accurate solutions to support tectonic motion modelling of the Azores region on a large scale, but that the recent stations still do not provide enough information to allow us to distinguish intra-island deformations due to their small relative magnitude.

In fact, Fig. 5.13 also shows that, independently of the tectonic model used, DEOSVel or NUVEL1-A, the relative deformations in the entire area of the plate boundary cannot be larger than 4–5 mm/year (at the islands it is expected to be at the millmetre/year level at the most).

Figure 5.14 shows an example of a model obtained from this kind of data (motions derived from geodetic observations). This model used a segment pattern deduced from magnetic data on the one hand and bathymetric and topographic features on the other in a multidisciplinary approach. The best configuration for the segment pattern was chosen by evaluating an elastic model which used the velocities provided by DEOSVEL as boundary conditions. One can see that Corvo on the North American plate is moving away from Graciosa and Santa Maria at a rate of 23.1 and 18.7 mm/year, respectively, which is due to the spreading of the plates at the mid-Atlantic ridge. In addition, one can also see that Graciosa and Santa Maria are separating from each other at a rate of 2.1 mm/year.

Permanent GPS data have been available in some of the Azores stations since 1999, allowing a more robust estimation of the secular motion at their locations. In order to complete the information from the limited number of available permanent stations, in our results we have incorporated solutions derived from periodic

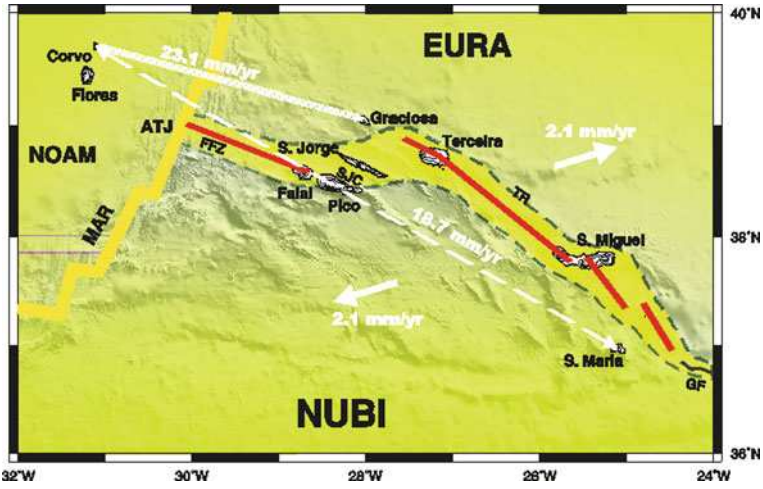


Fig. 5.14 Schematic model of the present-day kinematics of the Azores Triple Junction (Fernandes 2004)

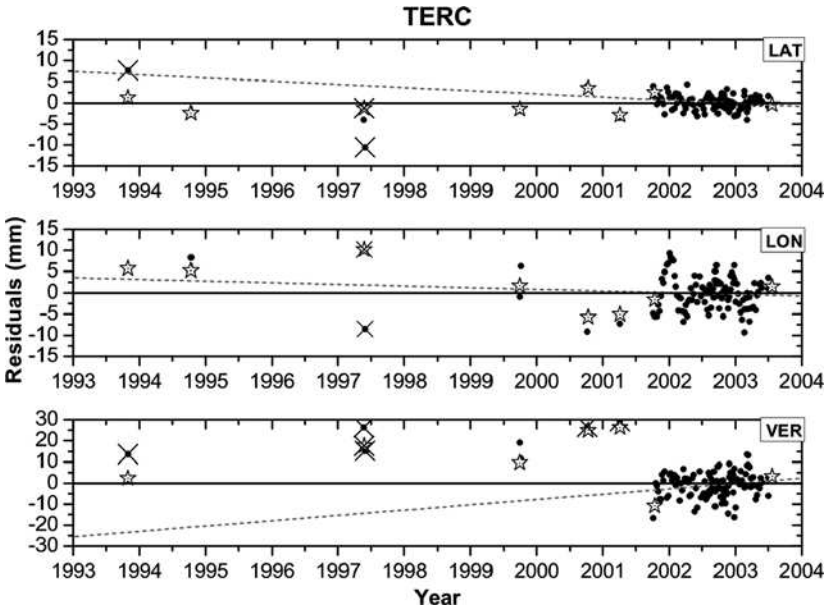


Fig. 5.15 Detrended time-series for TERC (permanent station since October 2001). Solutions considered outliers are crossed (Fernandes 2004)

observations that have been being carried out in the region since 1988 (Fernandes 2004).

After estimating the secular motion, one can subtract this motion from the observations and these detrended time-series for one of the TANGO stations (TERC) are given in Fig. 5.15. White stars show the residuals with respect to the estimated motion only with the TANGO campaign epochs. Dashed lines show the variation on the best-fit trend-line considering only the permanent data. Since the dashed lines differ from the horizontal, one can see that the derived velocity changed when changing to the permanent observation approach. This also allowed a significant improvement in the error assessment.

The campaign data provided reliable information to determine the tectonic motions due to the long time span of observations. The permanent GPS observations give more information about the seasonal variations in the position. To minimise these effects in data from episodic campaign the observations should be repeated at the same time of the year. In permanent data, such seasonal effects can be mitigated using estimated annual periodical signals. According to Blewitt and Lavallée (2002), the use of a minimum of 2.5 years of continuous observations is recommended in order to filter out the seasonal effects. We got an improvement in the confidence of the estimated motion using this approach.

The space-geodetic measurements can also be used to derive the geometry of an earthquake rupture. As an example, we show in Fig. 5.16 the displacement model computed for the Faial 1998 earthquake using the inverted coseismic displacements

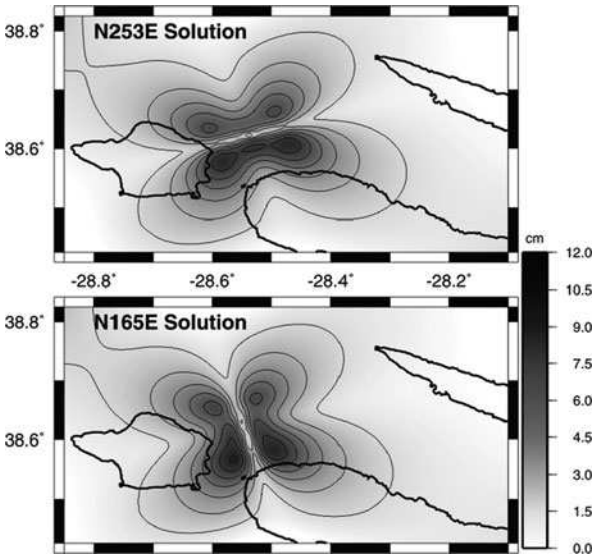


Fig. 5.16 Coseismic displacement fields at the Faial area. The solution for the N253E solution is shown *above*; the N165E solution is shown *below*

obtained from GPS. Two basic solutions were studied using the GPS solutions of 30 stations (Fernandes et al. 2002). Due to the location of the stations (all situated west of the epicentre), no absolute conclusion was drawn about the orientation of the rupture fault (left-lateral or right-lateral strike-slip). Nevertheless, the work showed the viability of such modelling, confirming that geodetic information can give a unique contribution to unravel the geometry of a fault rupture.

GPS has proven to be a valuable tool to study the present-day kinematics of plate boundaries, being particularly relevant in regions, such as the Azores, where geographical constraints (relative small percentage of emerged land) limit the design of an optimal spatial coverage demanding, and even more, the use of multidisciplinary approaches.

5.7 Further Developments

The development of the GNSS technology in the last two decades pulled up a most significant change in our way of studying the dynamics of the Earth's surface, with new and important findings leading to innovative aspects on the methodologies used and to a better comprehension of the phenomena involved.

The increased density of some existing GPS networks (e.g. in the USA, Europe and Asia) is providing unprecedented spatial and temporal sampling of crustal deformation. In spite of the evolution towards the permanent network approach, there are still a great number of GPS observations available, acquired in campaign mode

with episodic occupations of sites, which are of scientific interest. Furthermore, in several regions, the desired densification is only feasible with this type of observation methodology. Strategies for optimal combination of permanent and episodic data providing high temporal resolutions with spatially dense campaign should be developed.

While the accuracy of the GPS results is improving, new features which are not completely understood start to emerge. A good way to try to explain these features is through an intercomparison with other geodetic techniques. This can be either other space-geodetic techniques such as InSAR, SLR or VLBI or ground-based techniques such as levelling and absolute gravity measurements.

Terrestrial gravity observations differ from the other techniques since they provide information about the mass variations underneath the surface and not only the displacement of the Earth's surface. These observations are complemented by new space missions such as GRACE and GOCE that are bringing new insights concerning the Earth's gravity field, with impact on the geodetic and geodynamic applications.

Actual and new GNSS will provide data at sampling rates of 1 Hz and higher. This will especially make a major contribution to real-time geodynamic applications. However, in order to improve the accuracy of the measured high-rate displacements for applications such as seismology, it is important to reduce systematic errors at seismic frequencies. Developments are being made towards the use of high-rate GNSS to recover coseismic displacements. This will have a major impact on earthquake/volcano monitoring and tsunami early warning systems.

The use of GLONASS and of the future Galileo system for this type of applications will demand further developments in modelling the periodic signals. Special attention must be given to sub-daily effects present in the observations.

The spatial coverage of space-geodetic measurements is currently adding to the density needed to map the strain-rate distribution and to associate features in the deformation field with specific tectonic structures. These new data will allow very significant improvements in our understanding of coseismic, postseismic and interseismic deformation as well as substantial insights into the rheology of the crust and mantle.

GNSS will make a decisive contribution to the unravelling of earthquake generation processes and will undoubtedly be a strong basis for hazard risk assessments.

References

Agnew, D.C. (1992) The time-domain behavior of power-law noises. Geophys. Res. Lett., 19, 333–336

Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B. and Boucher, C. (2007) ITRF2005: a new release of the International Terrestrial Reference Frame based on time series of station positions and Earth Orientation Parameters. J. Geophys. Res., 112(B09401), doi: 10.1029/2007JB004949

Altamimi Z., Sillard, P. and Boucher, C. (2002) ITRF2000, a new release of the International Terrestrial Reference Frame for earth science applications. J. Geophys. Res., 107(B10), 2214, doi: 10.1029/2001JB000561

Bastos, L., Bos, M.S. and Fernandes, R.M.S. (2006) Investigation of common modes in the Iberian GPS network. Presented at the XIII Assembly of Wegener, Nice, 4–7 September 2006

Bastos, L., Bos, M.S., Fernandes, R.M.S. and Combrink, A. (2005) Impact of GPS time series errors in the estimation of tectonic plate models. Presented at the Joint Assembly of International Associations IAG, IAPSO and IABO, *Dynamic Planet* 2005, Cairns, 22–26 August 2005

Bastos, L., Osório, J., Barbeito, A. and Hein, G. (1998) Results from geodetic measurements in the western part of the African-Eurasian plate boundary. Tectonophysics, 294, 261–269

Bastos, L., Osório, J., Landau, H. and Hein, G. (1991) The Azores GPS Network, Arquipélago. Life Earth Sci., 9, 1–9

Beavan, J. (2005) Noise properties of continuous GPS data from concrete pillar geodetic monuments in New Zealand and comparison with data from US deep drilled braced monuments. J. Geophys. Res. 110(B8), CiteID B08410

Beran, J. (1994) "Statistics for Long-Memory Processes", Monographs on Statistics and Applied Probability, ISBN 0-412-04901-5

Beutler, G. (1996) The international GPS service for geodynamics: the story. Proceedings of the IAG Symposium, Vol. 115. Springer, Berlin, pp. 3–13

Bird, P. (2003) An updated digital model of plate boundaries. Geochem. Geophys. Geosyst., 4(3), 1027, doi: 10.1029/2001GC000252

Blewitt, G. (1993) Advances in Global Positioning System technology for geodynamics investigations: 1978–1992. In: Smith, D.E. and Turcotte, D.L. (eds.) Contributions of Space Geodesy to Geodynamics: Technolog, Geodyn. Ser., Vol. 25. Am. Geophys. Union. Washington, 195–213, 213 pp

Blewitt, G. and Lavallée, D. (2002) Effect of annual signals on geodetic velocity. J. Geophys. Res., 107(B7), 2145, doi: 1.1029/2001JB000570

Bock, Y., Agnew, D.C., Fang, P., Genrich, J.F., Hager, B.H., et al. (1993) Detection of crustal deformation from the Landers earthquake sequence using continuous geodetic measurements. Nature, 361(6410), 337–340

Bos, M.S., Fernandes, R.M.S., Williams, S.D.P. and Bastos, L. (2007) Fast error analysis of continuous GPS observations. J. Geod., doi: 10.1007/s00190-007-0165-x, online first

Bruyninx, C. (2000) Status and prospects of the permanent EUREF network. In: Gubler, E., et al. (eds) Proceedings of the Symposium of the IAG Subcommission for Europe, Vol. 8. EUREF Publication, Prague, June 2000, pp. 42–46

Camp, M. van, Williams, S.D.P. and Francis, O. (2005) Uncertainty of absolute gravity measurements. J. Geophys. Res., 110, B05406, doi: 10.1029/2004JB003497

Chase, C. (1978) Plate kinematics: the Americas, East Africa, and the rest of the world. Earth Planet Sci. Lett., 37, 355–368

Chavet, X, Valette, J.-J. and Feissel-Vernier, M. (2003) Analysis of geocenter time series derived form SLR, GPS and DORIS. AGU Fall Meeting, San Francisco, 8–12 December 2003

Degnan, J.J (1993) Millimeter accuracy satellite laser ranging: a review. In: Smith, D.E. and Turcotte, D.L. (eds.) AGU Monograph Geodynamics Series 25, Contributions of Space Geodesy to Geodynamics Technology, Washington, 133–162

DeMets, C., Gordon, R.G., Argus, D.F. and Stein, S. (1990) Current plate motions. Geophys. J. Int., 101, 425–478

DeMets, C., Gordon, R.G., Argus, D.F. and Stein, S. (1994) Effect of recent revisions to the geomagnetic reversal timescale on estimates of current plate motions. Geophys. Res. Lett., 21(20), 2191

Dong, D. and Bock, Y. (1989) Global Positioning System network analysis with phase ambiguity resolution applied to crustal deformation studies in California. J. Geophys. Res., 94, 3949–3966

EPNCB (2008) EUREF Permanent GPS Network – Data & Products – Products, retrieved January 15, 2008, http://epncb.oma.be/dataproducts/products/index.html

Feigl, K.L., Agnew, D.C., Bock, Y., Dong, D., Donnellan, A., et al. (1993) Space geodetic measurement of crustal deformation in central and southern California, 1984–1992. J. Geophys. Res., 98(B12), 1677–1712

Feissel-Vernier M, Le Bail, K., Berio, P., Coulot, D., Ramillien, G. and Valette, J.-J. (2006) Geocenter motion measured with DORIS and SLR, and geophysical evidence. J. Geodesy, 80(8–11)

Fernandes, R.M.S. (2004) Present-Day Kinematics at the Azores-Gibraltar Plate Boundary as Derived from GPS Observations. DUP, Delft, ISBN 90-407-2557-8

Fernandes, R.M.S., Ambrosius, B.A.C., Noomen, R., Bastos, L., Wortel, M.J.R., Spakman, W. and Govers, R. (2003) The relative motion between Africa and Eurasia as derived from ITRF2000 and GPS data. Geophys. Res. Lett., 30(16), 1828, doi: 10.1029/2003GL017089

Fernandes, R.M.S., Bastos, L., Ambrosius, B.A.C., Noomen, R., Matheussen, S. and Baptista, P. (2004) Recent geodetic results in the Azores Triple Junction region. Pure Appl. Geophys., 161(3), 683–699, doi: 10.1007/s00024–003–2469–y

Fernandes, R.M.S., Bastos, L., Miranda, J.M., Lourenço, N., Ambrosius, B.A.C., Noomen, R. and Simons, W. (2006) Defining the Plate Boundaries in the Azores Region. J. Volcanol. Geothermal Res., 156, 1–9, doi: 10.1016/j.jvolgeores.2006.03.0192006

Fernandes, R.M.S., Miranda, J.M., Catalão, J., Luis, J.F., Bastos, L. and Ambrosius, B.A.C. (2002) Coseismic displacements of the MW = 6.1, July 9, 1998, Faial earthquake (Azores, North Atlantic). Geophys. Res. Lett., 29(16), 1774, doi: 10.1029/2001GL014415

Gross, R., Fukumori, I. and Menemenlis, D. (2003) Atmospheric and oceanic excitation of the Earth's wobbles during 1980–2000. J. Geophys. Res., 108, 2370, doi: 10.1029/2002JB002143

Herring, T.A. (1999) Geodetic applications of GPS. Proc. IEEE, 87, 92–110

Hoffman-Wellenhof, B., Lichteneger, H. and Collins, J. (1997) Global Positioning System: Theory and Practice, 4th Ed. Springer, Berlin

Johnson, H. and Agnew, D. (1995) Monument motion and measurements of crustal velocities. Geophys. Res. Lett., 22, 2905–2908

Kahle, H.-G., Müller, M.V., Geiger, A., Danuser, G., Müller, S., Veis, G., Billiris, H. and Paradissis, D. (1995) The strain field in northwestern Greece and the Ionian Islands: results inferred from GPS measurements. Tectonophysics, 249(1–2), 41–52

Kasdin, N.J. (1995) Discrete simulation of colored noise and stochastic processes and 1/f power law noise generation. Proc. IEEE, 3(5), 802–827

Kellogg, J.N. and Dixon, T.H. (1990) Central and South America GPS geodesy – CASA UNO. Geophys. Res. Lett., 17(3), 195–198

Kreemer, C., Lavallée, D.A., Blewitt, G. and Holt, W.E. (2006) On the stability of a geodetic no-net-rotation frame and its implication for the International Terrestrial Reference Frame. Geophys. Res. Lett., 33, L17306, doi: 10.1029/2006GL027058

Langbein, J. (2004) Noise in two-color electronic distance meter measurements revisited. J. Geophys, Res., 109(B4), CiteID B04406

Larsen, S. and Reilinger, R. (1992) Global Positioning System measurements of strain accumulation across the Imperial Valley, California: 1986–1989. J. Geophys. Res., 97(B4), 8865–8876

Larson, K.M., Freymueller, J.T. and Philipsen, S. (1997) Global plate velocities from the Global Positioning System. J. Geophys. Res., 102(B5), 9961–9982

Mao, A., Harrison, C.G.A. and Dixon, T.H. (1999) Noise in GPS coordinate time series. J. Geophys. Res., 104(B2), 2797–2816

McCarthy, D. and Petit, G. (eds) (2004) IERS Conventions (2003). IERS Tech. Note 32, Verl. des Bundesamts fur Kartogr. und Geod., Frankfurt am Main, 127 pp

Minster, J.B. and Jordan, T.H. (1978) Present-day plate motions. J. Geophys. Res., 83(B11), 5331–5354

Miyazaki, S., Tsuji, H., Hatanaka, Y., Abe, Y., Yoshimura, A., Kamada, K., Kobayashi, K., Morishita, H. and Iimura, Y. (1996) Establishment of the nationwide GPS array (GRAPES) and its initial results on the crustal deformation of Japan. Bull. Geogr. Surv. Inst. (Jpn.), 42, 27–41

Navarro, A., Catalão, J., Miranda, J.M. and Fernandes, R.M.S. (2003) Estimation of the Terceira Island (Azores) main strain rates from GPS data. Earths, Planets Space, 55, 637–642

Noomen, R., et al. (1996) Crustal deformations in the Mediterranean area computed from SLR and GPS observations. J. Geodynamics, 21(1), 73–96

Norabuena, E., Leffler-Griffin, L., Mao, A., Dixon, T.H., Stein, S., Seth, Selwyn, S.I., Ocola, L. and Ellis, M. (1998) Space geodetic observations of Nazca–South America convergence across the Central Andes. Science, 279(5349), 358

Parkinson, B.W., Spilker J.J., Jr., Axelrad, P. and Enge, P. (eds) (1996) Global Positioning System: Theory and Applications. Vol. 2, American Institute of Aeronautics and Astronautics, Washington

Penna N.T., King, M.A. and Stewart, M.P. (2007) GPS height time series: short-period origins of spurious long-period signals. J. Geophys. Res., 112(B2), B02402

Pollitz, F.F. (2003) Transient rheology of the uppermost mantle beneath the Mojave Desert, California, Earth Planet. Sci. Lett., 215(1–2), 89–104, doi: 10.1016/S0012-821X(03)00432-1

Ray, R.D., Eanes, R.J. and Lemoine, F.G. (2001) Constraints on energy dissipation in the earth's body tide from satellite tracking and altimetry. Geophys. J. Inter., 144, 471–480

Reilinger, R.E., McClusky, S.C., Oral, M.B., King, R.W., Toksoz, M.N., Barka, A.A., Kinik, I., Lenk, O. and Sanli, I. (1997) Global Positioning System measurements of present-day crustal movements in the Arabia–Africa–Eurasia plate collision zone. J. Geophys. Res., 102(B5), 9983–9999

Ryan, J.W. and Ma, C. (1998) NASA-GSFC's geodetic VLBI program: a twenty-year retrospective. Phys. Chem. Earth, 23(9–10), 1041–1052, doi: 10.1016/S0079-1946(98)00144-X

Scherneck, H.G., Johansson, J.M., Koivula, H., van Dam, T. and Davis, J.L. (2003) Vertical crustal motion observed in the BIFROST project. J Geodyn., 35, 425–441

Segall, P. and Davis, J.L. (1997) GPS Applications for geodynamics and earthquake studies. Ann. Rev. Earth Planet. Sci., 25, 301–336, doi: 10.1146/annurev.earth.25.1.301

Sella, G.F., Dixon, T.H. and Mao, A (2002) REVEL: a model for recent plate velocities from space geodesy. J. Geophys. Res., 107(B4), 2081, doi: 10.1029/2000JB000033, 1–30

Shimada, S. and Bock, Y. (1992) Crustal deformation measurements in central Japan determined by a Global Positioning System fixed point network. J. Geophys. Res., 97, 12437–12455

Soudarin, L. and Grétaux, J.F. (2006) A model of present-day tectonic plate motions from 12 years of DORIS measurements. J. Geodesy., 80, 609–624, doi: 10.1007/s00190-006-0090-4

Stamps, S., Calais, E., Saria, E., Hartnady, C., Nocquet, J.M., Ebinger, C. and Fernandes, R.M.S. (2008) A Kinematic model for the east african rift. Geophys. Res. Lett., 35, L05304, doi: 10.1029/2007GL032781

Stewart, M.P., Penna, N.T. and Lichti, D.D. (2005) Investigating the propagation mechanism of unmodelled systematic errors on coordinate time series estimated using least squares. J. Geodesy, 79(8), 479–489, doi: 10.1007/s00190-005-0478-6

Straub, C. and Kahle, H.-G. (1995) Active crustal deformation in the Marmara Sea region, NW Anatolia, inferred from GPS measurements. Geophys. Res. Lett., 22, 2533–2536

Tsuji, H., Hatanaka, Y., Sagiya, T. and Hashimoto, M. (1995) Coseismic crustal deformation from the 1994 Hokkaido-Toho-Oki earthquake monitored by a nationwide continuous GPS array in Japan. Geophys. Res. Lett., 22(13), 1669–1672

Vermaat, E., Degnan, J.J., Dunn, P., Noomen, R. and Sinclair, A. (1998) Satellite laser ranging, status and impact for Wegener. J. Geodyn., 25, 195–212

Watson, C., Tregoning, P. and Coleman, R. (2006) The impact of solid Earth tide models on GPS time series analysis. Geophys. Res. Lett., 33(8), L08306, doi: 10.1029/2005GL025538

Wdowinski, S., Bock, Y., Zhang, J. and Fang, P. (1997) Southern California Permanent GPS geodetic array: spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. J. Geophys. Res., 102, 18057–18070

Williams, S.D.P. (2003a) The effect of coloured noise on the uncertainties of rates from geodetic time series. J. Geodesy, 76(9–10), 483–494, doi: 10.1007/s00190-002-0283-4

Williams, S.D.P. (2003b) Offsets in Global Positioning System time series. J. Geophys. Res., 108(B6), ETG 12-1–12-13

Williams, S.D.P (2008) CATS: GPS coordinate time series analysis software. GPS Solutions, doi: 10.1007/s10291-007-0086-4, online first

Williams S.D.P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R.M., Prawirodirdjo, L., Miller, M. and Johnson, D.J. (2004) Error analysis of continuous GPS position time series. J. Geophys. Res., 109, B03412, doi: 10.1029/2003JB002741

Zerbini S., Richter, B., Negusini, M., Romagnoli, C., Simon, D., Domenichini, F. and Schwahn, W. (2001) Height and gravity variations by continuous GPS, gravity and environmental parameter observations in the southern Po Plain, near Bologna, Italy. Earth Planet. Sci. Lett., 192(3), 267–279

Zhang J., Bock, Y., Johnson, H., Fang, P., Williams, S.D.P., Genrich, J., Wdowinski, S. and Behr, J. (1997) Southern California permanent GPS geodetic array: error analysis of daily position estimates and site velocities. J. Geophys. Res., 102(B8), 18035–18055

Chapter 6
Earth Rotation

Florian Seitz and Harald Schuh

Contents

The rotation of the Earth varies continuously. Its rotation axis changes its orientation with respect to both a space-fixed and an Earth-fixed reference system, and the angular velocity of the rotation fluctuates with time. The knowledge and therewith the continuous observation of Earth rotation variations is important for various reasons. It is fundamental for the realisation of time systems, the accurate determination of reference frames and precise navigation by providing the link between an Earth-fixed and a space-fixed coordinate system. Moreover, time series of Earth rotation parameters are of great interest for various disciplines of geosciences and astronomy since their changes are related to gravitational and geodynamic processes in the Earth system. In this way, Earth rotation monitoring contributes significantly to the understanding of the dynamics of the Earth system and the interactions between its individual components, e.g. the exchange of angular momentum between atmosphere, ocean and solid Earth, or the coupling mechanism between the Earth's core and mantle. Today the metrological basis for this highly interdisciplinary research is provided by precise space geodetic techniques such as Very Long

F. Seitz (✉)
Earth Oriented Space Science and Technology, Technische Universität München (TUM), Arcisstr. 21, D-80333 Munich, Munich, Germany
e-mail: seitz@bv.tum.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_6,

Baseline Interferometry (VLBI), Satellite/Lunar Laser Ranging (SLR/LLR), Global Navigation Satellite Systems (GNSS) and ring laser gyroscopes.

6.1 Reference Systems

Generally speaking the rotation of the Earth can be interpreted as a change of the orientation of an Earth-fixed reference system $\mathcal{H}$ relative to a space-fixed reference system $\mathcal{I}$.

The rotation vector of the Earth ω changes its orientation and its absolute value with respect to either system. Independent of the coordinate system, the rotation vector is the vector that provides the direction of the instantaneous rotation axis. Its absolute value equals the instantaneous angular velocity of Earth rotation. The temporal variations of the Earth rotation vector in the space-fixed reference system are known as precession and nutation. Both are caused by lunisolar gravitational torques which can be described as functions of time by series expansions with high accuracy. The effects of precession and nutation have been known for centuries from astronomical observations. The change of the direction of the Earth rotation vector with respect to an Earth-fixed reference system is referred to as polar motion and was not observed before the end of the nineteenth century. Different to precession and nutation, polar motion and the variation of the Earth's angular velocity are not easily predictable since they are affected by a multitude of irregular geodynamic processes.

According to a fundamental theorem of rotational dynamics, the temporal derivative of the rotation vector of a rotating body is equal with regard to a body-fixed and a space-fixed reference system. The temporal derivative $\frac{\mathrm{d}x}{\mathrm{d}t}$ of an arbitrary vector x with respect to a body-fixed system and its temporal derivative $\frac{\mathrm{D}x}{\mathrm{D}t}$ with respect to a space-fixed system are related by

$$\frac{\mathrm{D}x}{\mathrm{D}t} = \frac{\mathrm{d}x}{\mathrm{d}t} + \omega \times x \,. \tag{1}$$

If the Earth rotation vector ω is introduced instead of x, the equation turns into

$$\frac{\mathrm{D}\omega}{\mathrm{D}t} = \frac{\mathrm{d}\omega}{\mathrm{d}t} + \omega \times \omega = \frac{\mathrm{d}\omega}{\mathrm{d}t} \,. \tag{2}$$

The equality of the derivatives means that the derivative of both the orientation of the rotation vector and its absolute value is identical in the two systems. Consequently the variations of the orientation of the Earth rotation axis in the space-fixed and in the Earth-fixed reference system are not independent of each other. The relation between the coordinates of the Earth rotation vector with regard to a space-fixed or Earth-fixed system and the temporal derivatives of the orientation parameters are expressed by Euler's kinematical equations (Moritz and Mueller 1987).

Let $e_i^{\mathcal{H}}$ and $e_i^{\mathcal{I}}$ $(i = 1, 2, 3)$ be the orthonormal base vectors of the two above-mentioned reference systems. The orientation of the Earth-fixed system with respect to the space-fixed system can then be written as

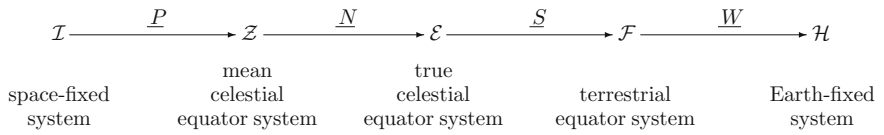
$$e_i^{\mathcal{H}} = \underline{R}\, e_i^{\mathcal{I}} \,, \tag{3}$$

where $\underline{R}$ means a time-dependent rotation matrix which is customarily composed of four parts (Richter 1995, McCarthy and Capitaine 2002):

$$\underline{R} = \underline{W}\, \underline{S}\, \underline{N}\, \underline{P} \,. \tag{4}$$

The matrices $\underline{P}$ and $\underline{N}$ stand for precession and nutation, respectively. The matrix $\underline{S} = R_3(\theta)$ is a spin at the so-called Earth rotation angle θ around the axis of the Celestial Intermediate Pole. $\underline{W}$ accounts for the components x and y of polar motion.

The transition from the space-fixed system $\mathcal{I}$ to the Earth-fixed system $\mathcal{H}$ is depicted as follows:



Today's fundamental astronomical space-fixed reference system is the *International Celestial Reference System* (ICRS) which was established by the International Astronomical Union (IAU) in 1997 (Feissel and Mignard 1998). The ICRS is a kinematically non-rotating coordinate system of high precision. Its origin is defined to be at the barycentre of the solar system. The ICRS which replaced the previous Fundamental Catalogue FK5 (Fricke et al. 1988) is realised in the radio frequency domain by the International Celestial Reference Frame (ICRF). The ICRF is described by equatorial coordinates of extragalactic and compact radio sources which are estimated from VLBI observations (Ma et al. 1998). At optical wavelengths the ICRS is realised by the Hipparcos catalogue. In 1998 the ICRF contained coordinates of 608 radio sources, and up to now 109 additional sources have been added by two extensions ICRF-Ext.1 and ICRF-Ext.2 (Fey et al. 2004; Gontier et al. 2006). A total of 212 very compact sources are used in order to define the axes of the reference frame (so-called defining sources). Presently the ICRF sources are observed with an accuracy of about 0.1 mas. VLBI is not capable of realising a geocentric ICRS, since it is a purely geometrical observation technique which does not provide any relation to the Earth's centre of mass. A Geocentric Celestial Reference Frame (GCRF) can be computed by combining VLBI and satellite observations or by referencing VLBI stations in a satellite-based geocentric reference frame (Seitz 2009). If the origin of the ICRS is shifted from the barycentre of the solar system into the Earth's centre of mass (under consideration of relativistic effects), the system experiences slight accelerations due to the motion of the Earth around the

Sun. Strictly speaking, such a Geocentric Celestial Reference System (GCRS) is no longer an inertial system. Commonly it is referred to as a *quasi-inertial system*.

As a consequence of its rotation the Earth is flattened at the poles. Since the Sun and Moon are generally located above or underneath the equatorial plane, a gravitational torque forces the equatorial plane towards the ecliptic (Torge 2001). Due to Earth rotation, this external force results in the precession of the Earth axis around the pole of the ecliptic, around which the rotation axis revolves on a cone with an apex angle of 23.5°. The vernal equinox that marks the intersection point of equatorial plane, ecliptic plane and the celestial sphere performs a clockwise motion at a rate of approximately 50.3″ per year along the ecliptic. In about 25,800 years, one so-called Platonic year, the vernal equinox performs one complete revolution around the celestial sphere. The precession matrix $\underline{P}$ describes the transition from the quasi-inertial GCRS into the *mean celestial equator system* $\mathcal{Z}$ (Capitaine et al. 2002; Rothacher 2002).

Precession is superposed by the lunisolar nutation, which causes variations of the Earth rotation axis in the mean celestial equator system. Lunisolar nutation is a consequence of the periodically changing positions of the Moon and Sun relative to the Earth. It is composed of various oscillations with different amplitudes and periods between few days and 18.6 years with respect to the space-fixed system (Mathews et al. 2002). The most prominent fraction of nutation is caused by the inclination of the lunar orbit by about 5° with respect to the ecliptic (Torge 2001). The orbital node, i.e. the intersection line of the lunar orbital plane and the ecliptic, moves with a period of 18.6 years along the ecliptic. As a consequence, the normal vector of the lunar orbital plane revolves along a cone around the ecliptic normal vector. The torque exerted by the Moon on the flattened Earth varies with the same period: it is maximum when the node of the lunar orbit coincides with the intersection line of equatorial plane and ecliptic and the Moon reaches its maximum declination of +28.5° or −28.5°. Further nutation terms are caused by the motion of the Moon and Sun between the northern and southern hemispheres. They feature periods of half a month and half a year, respectively (Torge 2001). With an apex angle of less than 10″, nutation is significantly smaller than precession. The nutation matrix $\underline{N}$ describes the transformation between the mean celestial equator system and the *true celestial equator system* $\mathcal{E}$.

The pole of the true celestial equator system is also known as the Celestial Intermediate Pole (CIP). According to resolution B1.7 adopted by the IAU in the year 2000 the CIP has superseded the previously used Celestial Ephemeris Pole (CEP) since 1 January 2003 (Capitaine 2002; McCarthy and Petit 2004). In pursuance of this IAU resolution, the CIP is defined as the axis with respect to which the Earth rotation angle is defined. The location of the CIP in the Earth-fixed reference system is provided by the International Earth Rotation and Reference Systems Service (IERS) on the basis of space geodetic observations and underlying models. The CIP is defined in such a way that it performs motions with periods longer than 2 days with respect to the space-fixed reference system. In the Earth-fixed system, retrograde motions of the CIP with frequencies between 0.5 and 1.5 cycles per sidereal day are allocated to nutation, whereas all other motions are interpreted as

polar motion. The change of the concept from the CEP to the CIP required the introduction of the revised model for precession and nutation IAU 2000A (Souchay et al. 1999; Mathews et al. 2002) according to resolution B1.6 of the IAU in the year 2000 (McCarthy and Petit 2004). A comprehensive overview of the IAU 2000 resolutions and their implications is given by Kaplan (2005).

If the Earth were solid and external torques were neglected, its instantaneous rotation axis would be directed towards the CIP. But in reality there is a small deflection between the CIP and the instantaneous rotation axis which is known as Oppolzer motion (Schödlbauer 2000; Capitaine 2004). As a consequence of precession and nutation, the Earth rotation axis changes its direction with respect to the space-fixed reference system as a function of time. Associated variations of right ascension and declination of fixed stars must be taken into account in astronomical observations from the Earth surface. The (true) latitude of a station, i.e. the angle between the true equatorial plane and the zenith of the station, is unaffected by precession and nutation. Matrices $\underline{P}$ and $\underline{N}$ can be modelled and predicted on the basis of lunar and solar ephemerides with high accuracy (Lieske et al. 1977; Wahr 1981; Seidelmann 1992). Small corrections to the current model (*celestial pole offsets*) are routinely published by the IERS on its internet site (http://www.iers.org). They account for model imperfections as well as for unpredictable geophysical signals such as the free core nutation or the quasi-annual oscillation of the S1 thermal tide (Dehant et al. 1999; Vondrak et al. 2005). Together with the precession–nutation model IAU 2000A, the celestial pole offsets allow for a precise computation of the location of the CIP in the space-fixed GCRF as illustrated in Fig. 6.1 (coordinates X and Y).

The transformation between $\mathcal{E}$ and the Earth-fixed system $\mathcal{H}$ is carried out on the basis of the so-called *Earth rotation parameters*. The rotation matrix $\underline{S}$ describes the diurnal rotation around the z-axis of the true celestial equator system. It is applied in order to transform between the true celestial equator system and the *terrestrial equator system* $\mathcal{F}$. Before 1 January 2003 the matrix $\underline{S}$ was related to the Greenwich Apparent Sidereal Time (GAST), i.e. the apparent hour angle of Greenwich with respect to the true vernal equinox. GAST is related to the Greenwich Mean Sidereal

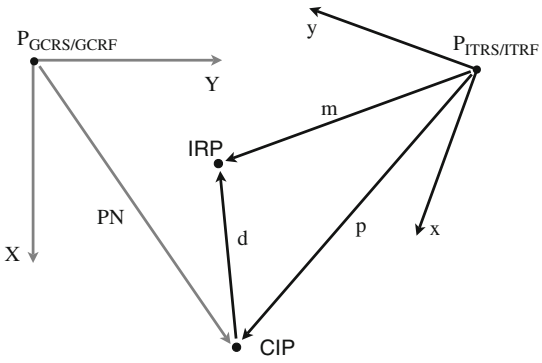


Fig. 6.1 Poles of reference with regard to the coordinate systems ITRS and GCRS (and their respective realisations ITRF/GCRF), and correspondence between model values $m(t)$ and published polar motion values $p(t)$ (Mendes Cerveira et al. 2009)

Time (GMST), i.e. the Greenwich hour angle of the mean vernal equinox, by the equation of equinoxes. From GMST universal time UT1 can be accessed (Aoki et al. 1982). According to the IAU resolution B1.8 (2000) and its supplement (IAU resolution B2, 2006), the vernal equinox as the direction of reference for the sidereal rotation of the Earth is now replaced by the so-called Celestial Intermediate Origin (CIO) in the space-fixed reference system (Capitaine 2002, 2008; McCarthy and Petit 2004). The CIO represents a non-rotating origin (Guinot 1979; Aoki and Kinoshita 1983) and is defined in such a way that the rotation vector of the celestial equator system with regard to a space-fixed reference system has no component in the direction of the CIP. The motion of the CIO relative to the space-fixed reference system has no component along the equator but a perpendicular one. Analogously a Terrestrial Intermediate Origin (TIO) is defined: the rotation vector of the terrestrial equator system with regard to an Earth-fixed reference system has no component in the direction of the CIP, and the motion of the TIO relative to the Earth-fixed reference system has solely a component perpendicular to the equator (Guinot 2002). In this concept GAST is replaced by the *Earth rotation angle* θ that is defined as the angle measured along the equator of the CIP between the unit vectors directed towards CIO and TIO. Since the direction of reference for UT1 moves uniformly along the equator, UT1 and θ are linearly related. The implementation of the IAU resolution B1.8 (2000) allows for a rigorous definition of the sidereal rotation of the Earth and for describing the rotation of the Earth independently from its orbital motion (McCarthy and Petit 2004).

The last part of the rotation matrix $\underline{R}$, the polar motion matrix $\underline{W}$, describes the transformation from the terrestrial equator system into the Earth-fixed system $\mathcal{H}$. The z-axis of the terrestrial equator system $\mathcal{F}$ is directed towards the CIP, while the z-axis of the terrestrial system is directed towards the Conventional Terrestrial Pole (CTP). Today the defined CTP is the *IERS Reference Pole*, which replaced the *Conventional International Origin* in the year 1967. The Conventional International Origin is identical with the mean direction of the Earth rotation axis between 1900 and 1905. The IERS Reference Pole differs from the Conventional International Origin by a maximum $\pm 0.03''$ and is realised by coordinates of globally distributed geodetic markers by means of space geodetic observations. Today's conventional Earth-fixed system $\mathcal{H}$ is the *International Terrestrial Reference System* (ITRS). Its origin is defined to be in the centre of mass of the Earth including atmosphere and ocean, and the z-axis of the right-hand system is directed towards the IERS Reference Pole. The orientation of the x-axis of the ITRS was originally defined by the Bureau International de l'Heure (BIH) for the epoch 1984.0. From this time, the evolution of the orientation was ensured by a no-net-rotation condition with regard to horizontal tectonic motions over the whole Earth (McCarthy and Petit 2004). The ITRS is realised by the determination of three-dimensional positions and velocities of geodetic observatories using space geodetic techniques. The most recent realisation of the ITRS is the ITRF2008. For details regarding the ITRF computation strategy see Altamimi et al. (2007).

Both the orientation of the rotation axis with respect to the CTP and the angular velocity of Earth rotation are influenced by transient, episodic and periodic exogenous and endogenous processes in the Earth system. Therefore the rotation

matrices $\underline{S}$ and $\underline{W}$ cannot be described or even predicted by models with satisfying accuracy. The IERS publishes different sets of Earth Orientation Parameters (EOP) in its circulars as well as on its internet site. Among the available parameters are the previously mentioned celestial pole offsets, the pole coordinates x_p and y_p and $\Delta\mathrm{UT} = \mathrm{UT1} - \mathrm{UTC}$. The pole coordinates x_p and y_p represent the misalignment between CIP and IERS Reference Pole, where the orientation of the x_p-axis is consistent with the x-axis of the ITRS, and the y_p-axis is directed towards 90° western longitude. The parameters x_p and y_p allow for the transformation between the terrestrial equator system $\mathcal{F}$ and the Earth-fixed system $\mathcal{H}$. Due to polar motion, the (true) latitude and longitude of a station on the Earth's surface vary with time.

Except for a constant offset due to the consideration of leap seconds, the coordinated universal time UTC corresponds to the uniform Temps Atomique International TAI which is realised by a set of worldwide distributed atomic clocks (BIPM 2007). Alternative to the parameter ΔUT, the expression *excess length-of-day* (ΔLOD) is common. ΔLOD is related to the absolute value of the Earth rotation vector in the terrestrial equator system and denotes the length of a solar day (length-of-day, LOD) expressed in UTC or TAI reduced by 86,400 s (Moritz and Mueller 1987):

$$\Delta\mathrm{LOD} = \mathrm{LOD} - 86,400\,\mathrm{s}\,. \tag{5}$$

ΔLOD and ΔUT are related according to

$$\Delta\mathrm{LOD} = -\frac{\mathrm{d}}{\mathrm{d}t}\Delta\mathrm{UT}\cdot 86,400\,\mathrm{s}\,. \tag{6}$$

Figuratively speaking, the term ΔLOD expresses the variation of the Earth's angular velocity due to geophysical and gravitational influences as a variation of the effective time for one full revolution. In former times ΔUT was observed by astronomical methods. Nowadays this parameter is unambiguously determined by VLBI due to its connection to the quasi-inertial reference frame of extragalactic radio sources. Global Navigation Satellite Systems (GNSS) allow for a precise observation of ΔLOD on short time scales.

6.2 Polar Motion

Figure 6.2 shows the Earth's polar motion between 1962 and 2009 as observed by astrometric and space geodetic observation techniques. The displayed values are taken from the well-known series EOP 05 C04 (Bizouard and Gambis 2009), in which the IERS publishes Earth orientation parameters together with respective formal errors at daily intervals since 1962. Values in this series are provided with respect to the precession–nutation model IAU 2000A and are consistent with the ITRF2005. Today polar motion can be determined with an accuracy of better than 0.1 mas (IERS 2008).

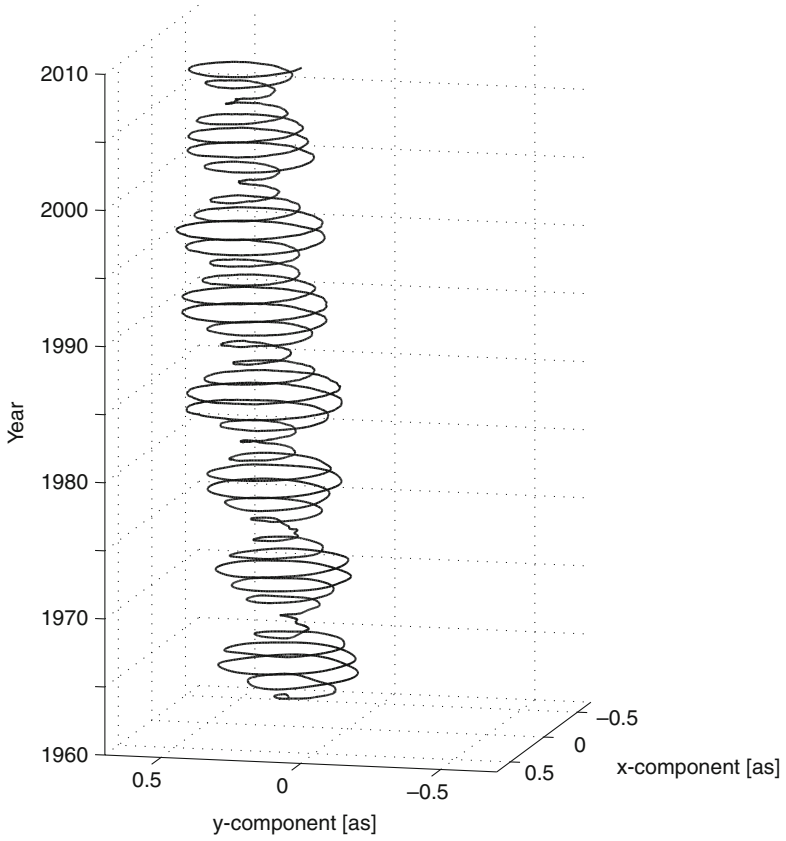


Fig. 6.2 Observations of polar motion from the EOP 05 C04 series of the IERS between 1962 and 2009

A clear beat with a period of 6.3 years is obvious. It is caused by the superposition of a signal component with annual period (approx mean amplitude 0.09″) and an oscillation with a period of about 1.2 years (approx mean amplitude 0.17″). The resulting beat amplitude is up to 0.25″ which corresponds to approximately 9 m on the Earth's surface.

While the annual oscillation can be explained by gravitational and geophysical effects within the Earth system, the oscillation with a period of 1.2 years is a free rotational mode of the Earth. It was discovered by Chandler (1891, 1892) and is therefore known as *Chandler oscillation*. The Chandler oscillation originates from a misalignment of the polar principal axis of inertia (figure axis) and the rotation axis of the Earth (Schödlbauer 2000). This causes a tumbling motion of the flattened Earth gyro, in which the rotation vector revolves on a cone around the figure axis. The Chandler oscillation is a prograde polar motion, i.e. counter-clockwise when seen from the North Pole. The existence of such a free oscillation of the Earth had

earlier been predicted by Euler (1765). From theoretical computations for a solid body with the Earth's dimension, he determined a period of 304 days (Euler period) for one revolution. Since the Earth is deformable, the actual period is lengthened to about 432 days (Chandler period) (see Sect. 6.4.2.1).

Signal decomposition of observed polar motion by means of wavelet filtering (Seitz and Schmidt 2005) allows for splitting the entire signal into its two main constituents, i.e. the Chandler oscillation and the annual oscillation. The resulting time series (x-components) are shown in Fig. 6.3 for a period of 150 years between 1860 and 2009. Since both signal components are almost circular, the y-components look very similar. Displayed values for polar motion are taken from the long-term C01 series, in which the IERS provides observations made since 1846 in a temporal resolution of 0.1 years (1846–1889) and 0.05 years (1890–2009). During the first decades the observations were based on optical astrometry and are comparatively inaccurate (standard deviations up to $\sigma = 0.16''$). The top panel of Fig. 6.3 shows the x-component of the time series C01 (after removal of a linear trend) together with the 3σ error margin. The Chandler oscillation (middle) features much stronger amplitude variations than the annual signal (bottom) which has been rather uniform during the last century (the first and the last years in the plot should not be interpreted due to boundary effects of the applied filter). Although the accuracy of the older astrometrical data is rather poor, the displayed amplitude variations are significant since the signal exceeds some 100 mas.

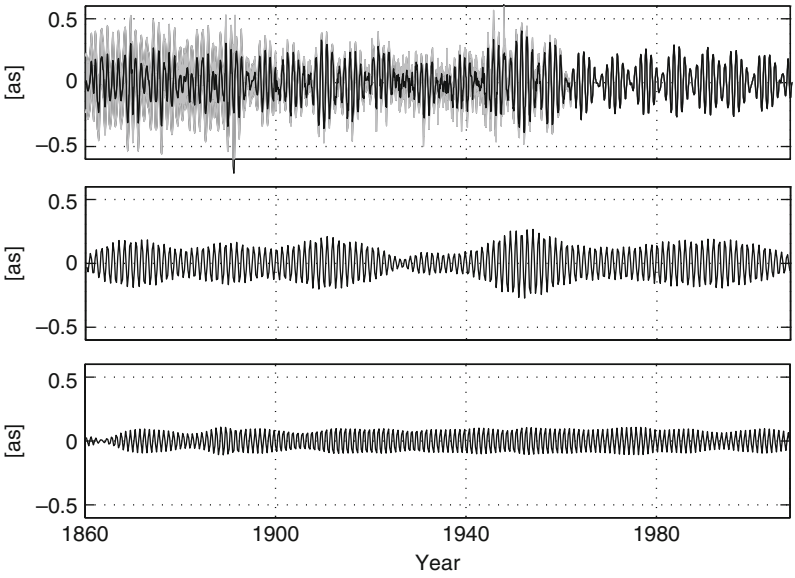


Fig. 6.3 Long-term observations of polar motion (x-component, linear trend removed) between 1860 and 2009 together with the 3σ error margin plotted in grey (*top*) and Chandler (*middle*) and annual (*bottom*) signal component determined by wavelet filtering

The origin of the strong amplitude variations and therewith the causative mechanism for the evocation of the Chandler oscillation have been under discussion for many years. As a consequence of the anelasticity of the Earth mantle and the associated dissipation due to friction, the Chandler oscillation is a damped oscillation. But the observations indicate that the amplitude of the free polar motion is excited by some mechanism which counteracts the damping. In numerous publications this matter has extensively been discussed. It has been investigated whether atmospheric or hydrologic mass redistributions (Wahr 1983; Hameed and Currie 1989; Sidorenkov 1992; Furuya et al. 1996, 1997) or processes in the Earth's interior (Souriau and Cazenave 1985; Gross 1986; Hinderer et al. 1987) are the hurriers of the oscillation. Since the Chandler oscillation is a resonance oscillation of the Earth, potential excitation mechanisms require energy in a band close to the Chandler frequency in order to excite the free polar motion and thus to counteract its damping. In recent years a number of studies came to the conclusion that the Chandler oscillation is excited by the combined effect of atmosphere and ocean (Gross 2000; Brzezinski and Nastula 2000; Seitz and Schmidt 2005). However, the individual contributions of these two subsystems could still not be fully assessed, since all investigations are naturally dependent on imperfect model assumptions of atmospheric and oceanic processes and their related mass transports. Furthermore, minor effects from continental hydrosphere, cryosphere and other subsystems must also be taken into account in order to close the budget of polar motion excitation.

The annual signal of polar motion originates similarly to a number of further significant higher and lower frequencies from gravitational and internal geophysical excitations, causing mass redistributions and mass motions within and between the Earth's subsystems. An overview of important drivers and the corresponding signatures in polar motion (amplitudes and periods) is given by Chao (1994) and Gross (2007). As mentioned above, there are also singular and non-periodic contributions from transient and episodic geophysical effects, such as earthquakes (Chao and Gross 1987, 2005) or El Niño situations (Kosek et al. 2001). Forced variations of polar motion and the free Chandler oscillation are closely linked. Variations of the Earth rotation vector induce a change of the Earth's centrifugal potential which leads to additional mass redistributions in the solid Earth and the ocean (so-called rotational deformations). This back-coupling effect causes a motion of the principal axis of inertia that affects the Chandler oscillation (see Sect. 6.4.2.1).

Figure 6.4 shows the polar motion curve in units of metres on the Earth surface in more detail for a time interval of 6 months (Schreiber et al. 2004). The large circle results from the superposition of signal components with comparatively long periods (especially the prograde Chandler and pro- and retrograde annual oscillations), whereas the small circles with magnitudes of approximately $0.01''$ are nearly diurnal retrograde polar motion components which are related to corresponding precession and dominant nutation terms in the space-fixed reference system (Oppolzer terms). The pronounced beat effect with a period of 13.7 days results from the superposition of oscillations that correspond to precession (period 0.997 days in the Earth-fixed reference system) and the largest nutation term (period 1.076 days in the Earth-fixed reference system). Variations of the beat amplitude are caused by further signal components that correspond to other nutation terms with approximately diurnal period

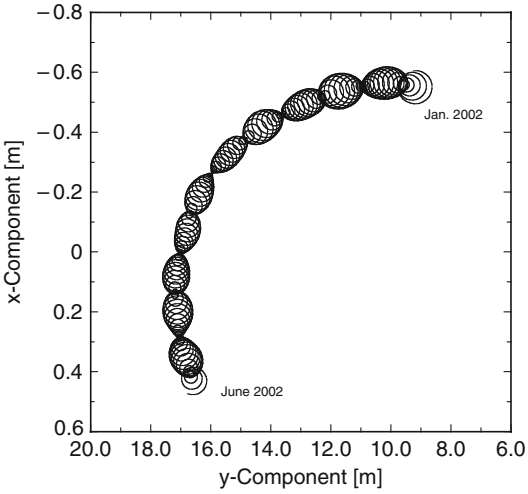


Fig. 6.4 Path of the rotation pole in the Earth-fixed reference system between January and June 2002 in metres on the Earth surface (Figure taken from Schreiber et al. 2004)

in the Earth-fixed system (McClure 1973). In analogy to precession and nutation in the space-fixed system, the retrograde nearly diurnal polar motion in the terrestrial system originates from lunisolar gravitational torques on the equatorial bulge of the Earth. Earth rotation causes a daily variation of the gravitational forces which results in the almost circular motion of the rotation pole in the direction opposite to the rotation. Nearly diurnal retrograde polar motion cannot be directly assessed by observations of VLBI, SLR/LLR and GNSS since these techniques are sensitive only to the complete rotation matrix from the Earth-fixed to the space-fixed reference frame from which no discrimination between celestial pole offsets and nearly diurnal retrograde polar motion is possible. An inertial rotation sensor on the Earth's surface is sensitive to the diurnal retrograde polar motion since the angle between the axis of the instrument and the rotation axis of the Earth changes with a period of 1 day. In this way, ring laser gyroscopes allow for the direct observation of the position of the instantaneous rotation axis and therewith for the assessment of the diurnal polar motion (Schreiber et al. 2004).

Beside the periodic and irregular fluctuations, polar motion is characterised by a secular trend at a present rate of 3.3 mas/a in the direction of 76°–78° western longitude (Vondrak et al. 1995; Schuh et al. 2001). Although the reason is not entirely understood yet, there is evidence that this secular motion is caused by postglacial rebound and sea-level variations (Milne and Mitrovica 1998).

6.3 Variations of Length-of-Day and ΔUT

The variation of the length of a solar day (ΔLOD) can be determined from the observations of modern space geodetic techniques with an accuracy of 20μs (IERS 2008). As shown in (6) ΔLOD is directly related to ΔUT. While accurate short-term time series of ΔLOD, i.e. of the derivative of ΔUT, can be estimated with high temporal resolution from GNSS observations, mid-term and long-term stability

of ΔLOD as well as ΔUT series can only be guaranteed by VLBI, providing the connection to the quasi-inertial reference frame. All satellite-based techniques, such as GPS or Glonass, meet the problem that Earth rotation cannot be distinguished from a uniform rotation of the satellite orbit nodes (Ray 1996).

Figure 6.5 (top panel) displays the observed variations of length-of-day from the EOP 05 C04 of the IERS for the period between 1962 and 2009. The curve is dominated by a secular signal of the order of milliseconds that is superposed by significant variations with annual and semi-annual periods due to mainly atmospheric effects and tidal signals with periods of several days. In contrast to polar motion, there is no free variation of length-of-day due to rotational deformations (Wahr 1985). The decadal variability of ΔLOD is ascribed to the exchange of angular momentum between the Earth's core and mantle (Liao and Greiner-Mai 1999). This assumption is supported by strong correlations between the decadal variations of ΔLOD with fluctuations of the Earth's magnetic field (Schuh et al. 2003). Four potential mechanisms of core–mantle coupling (CMC) are presently under discussion: topographic, electromagnetic, viscoelastic and gravitational coupling. Available models of topographic coupling are rather inaccurate since the knowledge of the topography at the core–mantle boundary is insufficient. But presumably this coupling mechanism does not provide enough energy in order to excite the strong variations of ΔLOD (Ponsar et al. 2002). Holme (1998) showed that the electromagnetic CMC seems to be the most important excitation mechanism. It is based on variations of the geomagnetic field due to dynamo processes, which exert a torque on conductive regions of the lower mantle via the Lorentz force (Schuh et al. 2003). Viscoelastic and gravitational coupling are inferior. In the frame of its *Special Bureau for the Core* (SBC) of the Global Geophysical Fluids Center (GGFC), the IERS provides model time series that describe the effects of CMC on ΔLOD. In Fig. 6.5b the results of three different models are compared with a moving average of the observations over 5 years. One of the model data sets (according to Jackson, Bloxham and Gubbins, JBG) has a temporal resolution of 1 year (Jackson 1997); the other two models (according to Petrov and Dehant, PD1 and PD2) are available for intervals of 5 years. All data sets are based on the frozen flux hypothesis (Jault et al. 1988). While JBG is a free model, PD1 and PD2 are based on observations of the magnetic field. The comparison of the various models provided by the SBC reveals significant differences. To a certain extent the data series correspond with the moving average (especially in the case of PD2), but the temporal resolution is much too coarse to explain the decadal variations of ΔLOD with sufficient accuracy and thus to exclude other causative processes.

Variations of ΔLOD on annual, seasonal and shorter time scales are highly correlated with angular momentum fluctuations within the atmosphere (mainly due to zonal winds) and, to a minor extent, due to ocean currents. The two strongest signal components induced by those processes, i.e. the annual and semi-annual oscillation, feature almost equal amplitudes of approximately 0.36 ms. In addition, there is a weak quasi-biennial oscillation (QBO) due to irregular variations of zonal winds and temperatures in the tropical troposphere and stratosphere (Trenberth 1980). Its amplitude varies from cycle to cycle. In general it is smaller than 0.1 ms (Höpfner

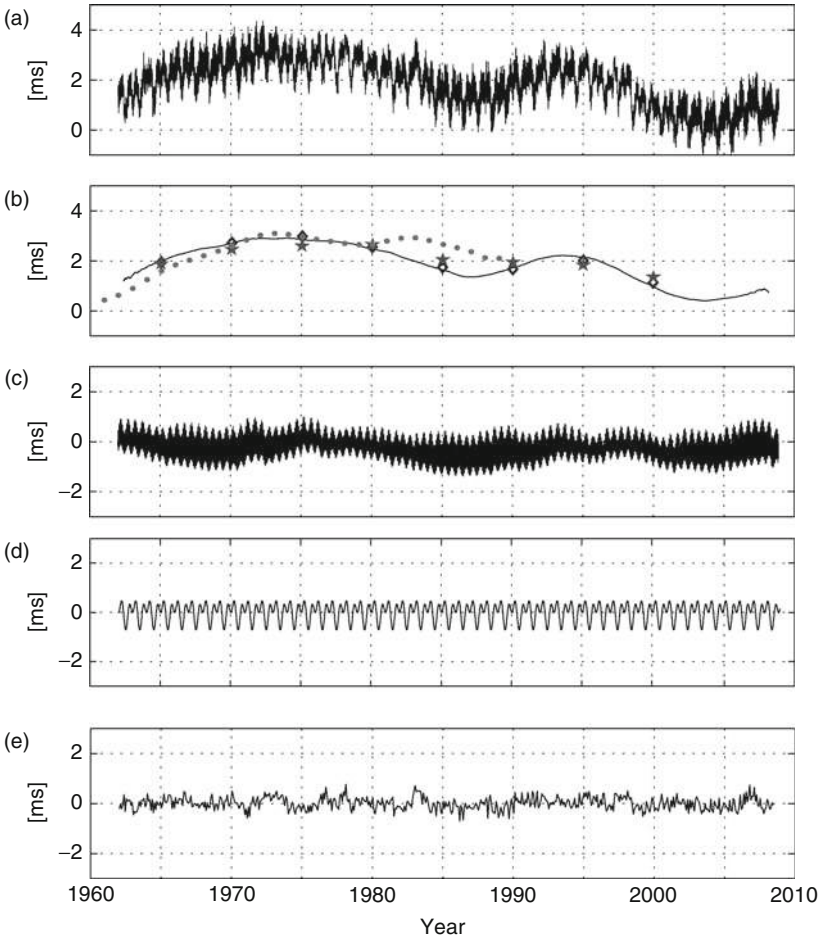


Fig. 6.5 Variations of length-of-day (ΔLOD) for the time frame between 1962 and 2009. (**a**) Observation time series EOP 05 C04. (**b**) Moving average over 5 years in comparison with three models for the influence of core–mantle interaction (dots: JBG; stars: PD1; diamonds: PD2; see text). (**c**) Effect of solid Earth tides. (**d**) Annual and semi-annual signal component. (**e**) Residual time series (**a-b-c-d**)

2001). The most important periods induced by solid Earth tides are 9.13 days (amplitude 0.07 ms), 13.63 days (0.15 ms), 13.66 days (0.35 ms) and 27.55 days (0.19 ms) (Yoder et al. 1981; McCarthy and Petit 2004). In contrast to solid Earth tides, the influence of ocean tides on ΔLOD is small (Lambeck 1980; Gross 1993), but not negligible in high-precision space geodesy.

The residual signal of ΔLOD (Fig. 6.5e), i.e. after reduction of the decadal signal, the annual and semi-annual oscillations and the tidal effects feature transient increases of the length-of-day during 1983 and (somewhat less pronounced) during 1997. These episodic signals can be explained by strong El Niño events (Rosen et al.

1984; Chao 1989). Like polar motion, ΔLOD is characterised by a secular change. Especially as a consequence of tidal friction, the length of a solar day increases by 2.3 ms per century (Morrison and Stephenson 1998).

6.4 Physical Model of Earth Rotation

6.4.1 Balance of Angular Momentum in the Earth System

From a physical perspective, Earth rotation can be interpreted as the rotary motion of a multitude of individual and interrelated mass elements about one common axis. This rotary motion is comparable to that of a physical gyroscope. Therefore theoretical and numerical studies on temporal variations of Earth rotation are based on equations of gyroscopic motion which follow from the balance of angular momentum in the Earth system. With respect to an Earth-fixed, i.e. rotating, reference system, the balance between the Earth's angular momentum $\mathbf{H}$ and external torques $\mathbf{L}$ due to, e.g., lunisolar and planetary gravitational forces is described by the dynamic Euler equation (Lambeck 1980):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{H} + \boldsymbol{\omega} \times \mathbf{H} = \mathbf{L} \,. \tag{7}$$

In this equation $\boldsymbol{\omega}$ denotes the rotation vector of the Earth with respect to the rotating reference system. The angular momentum of a rotating rigid body equals the product of its tensor of inertia $\underline{\mathbf{I}}$ and the rotation vector $\boldsymbol{\omega}$:

$$\mathbf{H} = \underline{\mathbf{I}} \cdot \boldsymbol{\omega} \,. \tag{8}$$

The symmetric tensor of inertia describes the mass distribution in the system (Lambeck 1980). In the case of a rigid body it is invariant with respect to body-fixed axes:

$$\underline{\mathbf{I}} = \int\int\int \rho(x,y,z) \begin{pmatrix} y^2+z^2 & -xy & -xz \\ -xy & x^2+z^2 & -yz \\ -xz & -yz & x^2+y^2 \end{pmatrix} \mathrm{dV} \,, \tag{9}$$

where $\rho(x, y, z)$ is the density at the three-dimensional position (x, y, z). In the case of a rotating deformable body, the angular momentum $\mathbf{H}$ is split into two parts: one fraction corresponds to the angular momentum of the rotating rigid body (8), but with the difference that the tensor of inertia is now time variable due to deformability. The second fraction can be viewed as angular momentum $\mathbf{h}$ relative to the body rotation. It follows from the motion of mass elements with velocity $\mathbf{v}^{\mathrm{rel}}$ relative to the rotating reference system, in which the rotation is described:

$$\mathbf{h} = \int\int\int \rho(x,y,z) \cdot (\mathbf{r} \times \mathbf{v}^{\mathrm{rel}})\, d\mathrm{V} \,, \tag{10}$$

where $\mathbf{r}$ denotes a three-dimensional position vector. Consequently the angular momentum of a rotating deformable body is (Schneider 1988)

$$\mathbf{H} = \underline{\mathbf{I}} \cdot \boldsymbol{\omega} + \mathbf{h} \,, \tag{11}$$

where the first summand is also referred to as *mass term*, the second one as *motion term*. Insertion of (11) into (7) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(\underline{\mathbf{I}} \cdot \boldsymbol{\omega} + \mathbf{h}) + \boldsymbol{\omega} \times (\underline{\mathbf{I}} \cdot \boldsymbol{\omega} + \mathbf{h}) = \mathbf{L} \,. \tag{12}$$

In this form the equation is also known as *Euler–Liouville* or in short *Liouville equation* (Munk and MacDonald 1960). In the context of Earth rotation studies, the term deformability not only refers to deformations of the Earth's body but also to mass redistributions within and between the various components of the Earth system. In particular, atmospheric and oceanic transport processes and related mass changes are very important on time scales from hours and days to several years. While the time-variable mass distribution in the system influences the tensor of inertia $\underline{\mathbf{I}}$, motions of mass elements with respect to the reference system cause relative angular momenta $\mathbf{h}$. Consequently all elements of the Liouville equation are time variable:

$$\underline{\mathbf{I}} = \underline{\mathbf{I}}(t), \ \mathbf{h} = \mathbf{h}(t), \ \boldsymbol{\omega} = \boldsymbol{\omega}(t), \ \mathbf{L} = \mathbf{L}(t) \,. \tag{13}$$

Angular momentum is exchanged among the individual components of the Earth system via mass transfer processes and torques. The occurrence of relative angular momenta is not necessarily linked to the appearance of variations of the tensor of inertia. Certainly most of the relevant processes influence both the mass and the motion term simultaneously. For instance, the atmospheric flow is generally related to variations of atmospheric pressure, and ocean circulation is usually accompanied by variations of ocean bottom pressure. But on the other hand mass motions are conceivable that do not influence the mass distribution in the Earth system and consequently the tensor of inertia. This is the case if one mass element is instantaneously replaced by a subsequent one (e.g. in a ring-like ocean current) or if the Earth's core experiences an acceleration with respect to the Earth's mantle. Vice versa vertical deformations of the Earth as a consequence of loading or the time-variable snow coverage could be mentioned as examples of mass redistributions without a significant influence on the motion term.

In theoretical studies on Earth rotation, the quantities in the Liouville equation are often related to a rotating reference system, according to which the mass elements of a rotating rigid body are invariant with respect to their position at all times. For a deformable Earth such a system can be defined by a minimum condition (Schneider 1988). An example is the Tisserand system (Tisserand 1891), for which the integral effect of the relative motions of mass elements with respect to the reference system is minimised ($\mathbf{h} = 0$). The application of the Tisserand system simplifies the Liouville equation (12) considerably. But on the other hand the definition of the

Tisserand system is hypothetical, since relative angular momenta (especially in the Earth's interior) are not accessible from observations on the Earth's surface (Engels and Grafarend 1999).

Numerical investigations are commonly performed in a geocentric terrestrial reference system. Its rotation axis is oriented towards the polar moment of inertia C of the Earth, its x-axis is directed towards the Greenwich meridian and its y-axis towards 90°E. The terrestrial system performs a uniform rotation about its z-axis with angular velocity $\Omega = 2\pi/86,164\,\mathrm{s}$. Temporal variations of the instantaneous Earth rotation vector $\boldsymbol{\omega}(t)$ are viewed as small deviations of the uniform rotation. In coordinates of the terrestrial system the Earth rotation vector is expressed as (Munk and MacDonald 1960)

$$\boldsymbol{\omega}(t) = \Omega \cdot \begin{pmatrix} m_1(t) \\ m_2(t) \\ 1 + m_3(t) \end{pmatrix}, \qquad m_i \ll 1 \,. \tag{14}$$

The dimensionless quantities $m_i(t)\,(i = 1, 2, 3)$ represent slight disturbances of the uniform rotation (Munk and MacDonald 1960). The two components $m_1(t)$ and $m_2(t)$ describe the time-variable orientation of the rotation axis with respect to the z-axis of the terrestrial system (polar motion). Deviations of the Earth's angular velocity with respect to Ω are associated with changes of the length-of-day. They follow from the temporal variation of the absolute value of the Earth rotation vector $|\boldsymbol{\omega}(t)|$ (Lambeck 1980; Schneider 1988):

$$|\boldsymbol{\omega}(t)| = \Omega \sqrt{m_1(t)^2 \; + m_2(t)^2 \; + (1 + m_3(t))^2} \quad \approx \quad \Omega\,(1 + m_3(t)) \,. \tag{15}$$

The error of ΔLOD due to this approximation is 10^{-16} s and therefore negligible.

The Earth's tensor of inertia $\underline{\mathbf{I}}(t)$ can be interpreted as the sum of two components $\underline{\mathbf{I}}_0$ and $\Delta\underline{\mathbf{I}}(t)$ (Lambeck 1980), where $\underline{\mathbf{I}}_0$ is an approximate tensor. If the axes of the reference system coincide with the principal axes of inertia, the approximate tensor has a diagonal structure:

$$\underline{\mathbf{I}}_0 = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} \,, \tag{16}$$

where A and B are the equatorial principal moments of inertia and C is the axial principal moment of inertia of the Earth ($C > B > A$). But the axes of the principal moments of inertia differ from the axes of the previously described terrestrial reference system by approximately 15° in the equatorial plane (Marchenko and Schwintzer 2003). This divergence has to be taken into account by means of a rotation. Consequently $\underline{\mathbf{I}}_0$ does not have a diagonal structure with respect to the axes of the applied terrestrial system.

Due to mass redistributions in the Earth system, small time-dependent deviations $\Delta\underline{\mathbf{I}}(t)$ of the approximate tensor $\underline{\mathbf{I}}_0$ arise (Moritz and Mueller 1987). With the

tensor elements (so-called deviation moments) $c_{ij}(t) \ll A, B, C\,(i,j = 1,2,3)$ the symmetric tensor $\Delta\underline{\mathbf{I}}(t)$ reads

$$\Delta\underline{\mathbf{I}}(t) = \begin{pmatrix} c_{11}(t) & c_{12}(t) & c_{13}(t) \\ & c_{22}(t) & c_{23}(t) \\ \text{sym.} & & c_{33}(t) \end{pmatrix} . \tag{17}$$

If deviations of the tensor $c_{ij}(t)$, relative angular momenta $\mathbf{h}(t)$ and external torques $\mathbf{L}(t)$ are provided from models or observations, the solution of the Liouville Equation for $\boldsymbol{\omega}(t)$ allows for the forward computation of Earth rotation variations. The relation between modelled values $m_i(t)$ and geodetic observations will be discussed in Sect. 6.5.

Two different approaches, the angular momentum approach and the torque approach, are in principle applicable for the set-up and solution of the Liouville equation. Theoretically both approaches are equivalent, but they differ conceptually with respect to their view of the Earth system. Accordingly, the procedures of modelling effects of the Earth's fluid components (e.g. atmosphere, ocean, continental hydrosphere) on Earth rotation are different (De Viron et al. 2005).

6.4.1.1 Angular Momentum Approach

The angular momentum approach is the classical approach for modelling Earth rotation. It has been described in various publications (Munk and MacDonald 1960; Lambeck 1980; Barnes et al. 1983; Moritz and Mueller 1987). The rotating body for which the Liouville equation is set up comprehends the solid Earth, atmosphere, hydrosphere and all other subsystems. In the absence of external lunisolar and (much smaller) planetary torques, this system of mass elements is viewed to be isolated, i.e. the right-hand side of (12) is zero, and the total angular momentum of the rotating body is conserved. Fractions of angular momentum can be transferred between the individual system components by redistributions and motions of masses. Changes of the angular momentum due to atmospheric, oceanic and other dynamic processes are associated with an opposite change of angular momentum of the solid Earth which is accompanied by variations of the Earth rotation vector $\boldsymbol{\omega}(t)$.

In the angular momentum approach, solely gravitational torques from external celestial bodies act on the rotating Earth. If the Sun, Moon and planets are viewed as point masses, the gravitational torque $\mathbf{L}(t)$ on the right-hand side of the Liouville equation (12) can be written as (Moritz and Mueller 1987; Beutler 2005)

$$\mathbf{L}(t) = \sum_j \frac{3GM_j}{r_{ej}^5(t)} \begin{pmatrix} y_j(t)\, z_j(t)\, (C - B) \\ x_j(t)\, z_j(t)\, (A - C) \\ x_j(t)\, y_j(t)\, (B - A) \end{pmatrix} . \tag{18}$$

In this equation G is the gravitational constant, and index j stands for the respective celestial body with the (point-)mass M_j; its geocentric distance is denoted with $r_{ej}(t); x_j(t), y_j(t), z_j(t)$ are its coordinates in the rotating reference system. In

its conventions the IERS recommends the use of the solar, lunar and planetary *JPL Development Ephemeris* DE405/LE405 (Standish 1998; McCarthy and Petit 2004).

Each relocation of mass elements within the system leads to an instantaneous change of the tensor of inertia $\Delta\underline{\mathbf{I}}(t)$. Deviation moments $c_{ij}(t)$ for the solid Earth result from deformations of the Earth's body as reaction to a tide generating potential, rotational variations and surface mass loads (Moritz and Mueller 1987; Seitz et al. 2004) (see Sect. 6.4.2). Relative angular momenta $\mathbf{h}(t)$ are due to the motion of individual mass elements relative to the terrestrial reference system.

The angular momentum approach corresponds to an abstract balance of angular momentum of all subsystems. Their individual contributions to the angular momentum budget are linearly superposed:

$$\begin{aligned} \underline{\mathbf{I}}(t) &= \underline{\mathbf{I}}_0 + \Delta\underline{\mathbf{I}}_{\text{ solid Earth}}(t) + \Delta\underline{\mathbf{I}}_{\text{ atmosphere}}(t) + \Delta\underline{\mathbf{I}}_{\text{ ocean}}(t) + \cdots, \\ \mathbf{h}(t) &= \mathbf{h}_{\text{ solid Earth}}(t) + \mathbf{h}_{\text{ atmosphere}}(t) + \mathbf{h}_{\text{ ocean}}(t) + \cdots . \end{aligned} \tag{19}$$

Variations of the tensor of inertia can be computed from modelled or observation-based mass balances of the Earth's subsystems. Relative angular momenta are derived from fluxes from global atmosphere and ocean circulation models.

6.4.1.2 Torque Approach

In the torque approach the effects of the Earth's fluid components, atmosphere and ocean, on the balance of angular momentum are modelled as (quasi-)external torques (Wahr 1982). That is, the integral effect of direct atmospheric and oceanic forces on the solid Earth appears in the vector $\mathbf{L}(t)$ on the right-hand side of the Liouville equation (12). Similar to the angular momentum approach, variations of the tensor of inertia $\Delta\underline{\mathbf{I}}(t)$ are due to deformations of the solid Earth caused by tides, surface mass loads and rotational variations. Since atmosphere and ocean are viewed as external systems, their mass redistributions do not affect the tensor of inertia. Likewise there are no relative angular momenta $\mathbf{h}(t)$ due to atmospheric and oceanic currents.

Torques between atmosphere/ocean and the solid Earth are assessed on the basis of global atmosphere and ocean circulation models. The acting torque is composed of three parts: pressure torque, gravitational torque and friction torque (De Viron et al. 2001). The pressure torque acts on the Earth's topography. It is derived from fields of surface and ocean bottom pressure and the gradient of the topography. The gravitational torque is a result of the interaction between the mass distributions within atmosphere/ocean and the solid Earth. The friction torque results from the relative motion of atmosphere and ocean currents with respect to the Earth surface. Since the friction drag of the Earth's surface is widely unknown it is particularly difficult to model (De Viron and Dehant 2003a). In a study on the influence of the atmospheric torque on polar motion De Viron et al. (1999) demonstrated that the time derivatives of the equatorial atmospheric angular momentum and the sum of the atmospheric equatorial torques agree well in the spectral range of longer than 1 day. Furthermore this study revealed that the magnitude of the equatorial components of

pressure and gravitational torque are almost equal (but with opposite signs) and that both contribute significantly stronger to polar motion than the friction torque.

The effects of atmospheric and oceanic pressure torque, gravitational torque and friction torque are superposed to the previously described external gravitational torque exerted by Sun, Moon and planets (18). Therefore the total torque $\mathbf{L}(t)$ can be written as

$$\mathbf{L}(t) = \mathbf{L}_{\text{pressure}}(t) + \mathbf{L}_{\text{gravitation}}(t) + \mathbf{L}_{\text{friction}}(t) + \mathbf{L}_{\text{external}}(t) \; . \tag{20}$$

Since lunisolar and planetary torques have a discrete spectrum in narrow vicinity of the diurnal retrograde frequency, they can be modelled quite well via harmonic expansion. Atmospheric and non-tidal oceanic torques, however, have a continuous spectrum and are thus unpredictable. Consequently, the modelling has to be performed in the time domain.

From the viewpoint of physical understanding, the torque approach is superior to the angular momentum approach. By modelling explicit interactions between atmosphere/ocean and the solid Earth via particular forces, it is possible to tell which specific processes lead to a change of the angular momentum budget and thus cause variations of Earth rotation. The torque approach is ideal for geographical studies since it allows for a direct identification of regions in which the interaction between atmosphere, ocean and the solid Earth is stronger than in others (De Viron and Dehant 2003b). In this way, the approach provides valuable physical insights into dynamic interactions in the Earth system.

The largest limitation for the torque approach is the lack of sufficiently accurate numerical models for the computation of the torques due to atmospheric and oceanic pressure, gravitation and friction. While model errors are not so crucial in the case of the angular momentum approach (where the errors smooth out due to the computation of one global value), the torque approach is highly sensitive to errors (De Viron and Dehant 2003b). As stated above, many of the parameters which are necessary for the computation of torques are not well known, e.g. the friction drag between air and Earth surface or between water and ocean bottom. Furthermore, the computation of the pressure torque is unsatisfactory due to the comparatively coarse spatial resolution of available orography models (De Viron et al. 1999; Stuck 2002).

Due to these data problems, atmospheric and oceanic angular momentum values presently appear to be more reliable for the interpretation of geodetic observations of Earth rotation. Nevertheless the torque approach is promising in the light of future model advancements.

6.4.2 Solid Earth Deformations

Mass redistributions and corresponding variations of the tensor of inertia are also caused by deformations of the solid Earth as a consequence of its reaction to the lunisolar and planetary tide generating potential, variations of the centrifugal potential due to polar motion and mass loads on the Earth surface.

Theoretical considerations on the effects of solid Earth and ocean tides on Earth rotation are provided together with elaborate instructions for numerical computations in the conventions of the IERS (McCarthy and Petit 2004). For particulars the reader is referred to this publication and the references therein. The following section will focus on the deformations induced by rotational variations and surface mass loads.

6.4.2.1 Rotational Deformations

Temporal variations of the rotation vector $\boldsymbol{\omega}(t)$ lead to variations of the Earth's centrifugal potential. This causes deformations of the solid Earth and the ocean which are also known as *rotational deformations*. While vertical deformations due to variations of the angular velocity of the rotation are below 0.5 mm at the Earth surface (Wahr 1985) and therefore negligible, the effects due to polar motion are up to 25 mm (Gipson and Ma 1998). These changes of the Earth's geometry are accompanied by variations of the tensor of inertia that are superposed to other deviations $c_{ij}(t)$ $(i,j = 1,2,3)$ due to mass redistributions induced by gravity and other geophysical effects. The back coupling from polar motion to the tensor of inertia influences the Earth's rotational dynamics significantly: it is well known that rotational deformations are responsible for the prolongation of the Euler period of 304 days (which is the period of the free oscillation of a rigid body with the Earth's dimensions) to the observed period of the free oscillation of about 432 days (Chandler period) (Moritz and Mueller 1987).

The effect of polar motion on the Earth's centrifugal potential is referred to as *pole tide*. Parameters $m_1(t)$ and $m_2(t)$ of the Earth rotation vector are related to temporal variations of the coefficients $\Delta C_{21}(t)$ and $\Delta S_{21}(t)$ of the spherical harmonic expansion of the geopotential (McCarthy and Petit 2004):

$$\begin{aligned} \Delta C_{21}(t) &= -\frac{\Omega^2 a^3}{3GM_\mathrm{E}} \left(\Re(k_2) \cdot m_1(t) + \Im(k_2) \cdot m_2(t) \right) , \\ \Delta S_{21}(t) &= -\frac{\Omega^2 a^3}{3GM_\mathrm{E}} \left(\Re(k_2) \cdot m_2(t) - \Im(k_2) \cdot m_1(t) \right) , \end{aligned} \tag{21}$$

where a and M_E stand for mean equatorial radius and total mass of the Earth. The effect of polar motion on rotational deformations and therewith on the variation of the geopotential depends on the Earth's rheological properties. In (21) the rheology is described by the complex pole tide Love number $k_2 = \Re(k_2) + i\Im(k_2)$, where $\Re$ and $\Im$ stand for real part and imaginary part, respectively.

The coefficients $\Delta C_{21}(t)$ and $\Delta S_{21}(t)$ are directly linked to the elements of $c_{13}(t$ and $c_{23}(t)$ of the tensor of inertia (Lambeck 1980):

$$\begin{aligned} \Delta C_{21}(t) &= \frac{-c_{13}(t)}{a^2 M_\mathrm{E}} , \\ \Delta S_{21}(t) &= \frac{-c_{23}(t)}{a^2 M_\mathrm{E}} . \end{aligned} \tag{22}$$

If the Earth was a rigid body, i.e. if the tensor of inertia was invariant with respect to time and there were no relative angular momenta, the Earth would rotate freely at the Euler period of 304 days as stated above. In an extensive study Smith and Dahlen (1981) discussed the consequences of deformability for the period of the free polar motion and derived an appropriate numerical value of the pole tide Love number k_2 in the light of mantle anelasticity and the dynamics of core and ocean. In a first step Smith and Dahlen (1981) approximated the Earth as a purely elastic body and neglected the dynamic response of core and ocean. The pole tide Love number was introduced with the (preliminary) numerical value of $k_2^* = 0.30088$, which was computed from the hydrostatic ellipsoidal Earth model 1066A (Gilbert and Dziewonski 1975). It was shown that the period of the free rotation of a fully elastic Earth would amount to 447 days, i.e. 143 days longer than that of a rigid body.

In order to refine the Earth's reaction on rotational variations, the effects of the dynamic fluid core, the equilibrium ocean pole tides and the mantle anelasticity must be taken into account for the computation of rotational deformations. In the following a simple Earth model will be discussed which consists of an anelastic mantle and a spherical liquid core. Both are assumed to be completely decoupled. Basic considerations on the application of such a model body for studies on Earth rotation can be found in, e.g. Moritz and Mueller (1987) and Brzezinski (2001). It is similar to the models introduced by Molodensky (1961) and Sasao et al. (1980), but in contrast to the latter studies, the approach does not account for the exchange of angular momentum between core and mantle. While the effects of core–mantle coupling on polar motion are significant mainly on subdaily time scales, there are huge decadal variations of ΔLOD due to the interaction of core and mantle (see Sect. 6.3). As a consequence of the decoupling, the principal moments of inertia A, B and C which are the parameters of the approximate tensor of inertia $\underline{\mathbf{I}}_0$ (16) of the entire Earth have to be replaced by A_m, B_m and C_m, which are attributed to the mantle alone. Since the core is assumed to be spherical, the principal moments of inertia used for the computation are derived from $A_\mathrm{m} = A - A_\mathrm{c}, B_\mathrm{m} = B - A_\mathrm{c}$, and $C_\mathrm{m} = C - A_\mathrm{c}$, where A_c denotes the principal moment of inertia of the spherical core. Its value is derived from (Sasao et al. 1980)

$$A_\mathrm{c} = A\frac{\xi}{\gamma}, \tag{23}$$

where ξ and γ are constants accounting for the rheology of mantle and core. The values provided by Sasao et al. (1980) are $\xi = 2.300 \times 10^{-4}$ and $\gamma = 1.970 \times 10^{-3}$. In a later study, Mathews et al. (1991) computed $\xi = 2.222 \times 10^{-4}$ and $\gamma = 1.965 \times 10^{-3}$ from the Preliminary Reference Earth Model PREM (Dziewonski and Anderson 1981). The non-participation of the core in the rotation shortens the period of the free polar motion by approximately 50.5 days (Smith and Dahlen 1981). That is, the period of the free rotation of a fully elastic Earth with liquid core would be around 396 days.

In order to account for the effects of ocean pole tides and mantle anelasticity, surcharges to the above given value for the elastic pole tide Love number k_2^* are

added (Smith and Dahlen 1981). The effective pole tide Love number k_2 becomes

$$k_2 = k_2^* + \Delta k_2^{\mathrm{O}} + \Delta k_2^{\mathrm{A}}\,, \tag{24}$$

where Δk_2^{O} and Δk_2^{A} denote the incremental corrections of the elastic pole tide Love number due to ocean pole tides and the anelastic response of the Earth's mantle. Following Smith and Dahlen (1981) and a more recent study by Mathews et al. (2002) the appropriate addend for the contribution of equilibrium ocean pole tides amounts to $\Delta k_2^{\mathrm{O}} = 0.044$. Thereby the period of the free oscillation is lengthened by about 29.8 days (Smith and Dahlen 1981).

The reaction of the Earth's mantle on variations of the centrifugal potential is not ideally elastic. Due to friction, rotational deformations of the mantle are a dissipative process which is equivalent to an attenuation of the free polar motion. That means, in the absence of a counteracting excitation mechanism, the rotation axis of the Earth would dislocate towards its figure axis within a few decades (Moritz and Mueller 1987). The effect of mantle anelasticity causes an extension of the period of the free rotation by another 8.5 days (Wilson and Haubrich 1976). It is considered by the complex surcharge $\Delta k_2^{\mathrm{A}} = 0.0125 + 0.0036i$ to the Love number k_2^* (Mathews et al. 2002; McCarthy and Petit 2004).

Summing up the effects of ocean, core and mantle, the value of the pole tide Love number is $k_2 = 0.35 + 0.0036i$ (McCarthy and Petit 2004). This value is appropriate for a deformable Earth with a spherical liquid core, taking into account the effects of ocean pole tides and mantle anelasticity. When k_2 was applied in a numerical simulation with a dynamic Earth system model, the resulting Chandler period was 431.9 days (Seitz et al. 2004) which coincides with geodetic observations. The result of the simulation for the x-component of polar motion over a period of 100 years is displayed in Fig. 6.6 Since neither gravitational effects nor mass redistributions and motions in the Earth's fluid components have been considered in this experiment, the curve reflects the free polar motion under the influence of mantle anelasticity or – mathematically speaking – under the influence of the imaginary part of the pole tide Love number $\Im(k_2)$. The curve is provided in normalised representation since the choice of the initial values is arbitrary. The damping of the Chandler amplitude is obvious, and after already 22 years the amplitude is reduced by half.

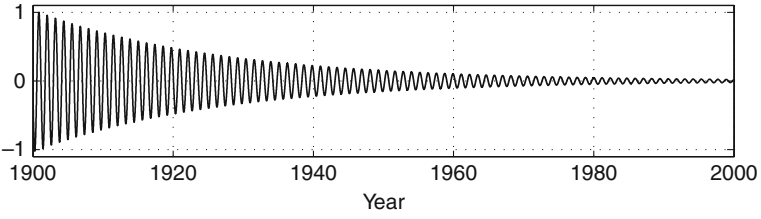


Fig. 6.6 Damped Chandler oscillation (x-component) derived from a simulation study with a dynamic Earth system model over 100 years regarding ocean pole tides and mantle anelasticity (Seitz et al. 2004)

The damping function $c(t)$ is the envelope of the oscillation

$$c(t) = c_0 \cdot e^{-\delta(t-t_0)}, \tag{25}$$

where c_0 is the initial amplitude of the oscillation and δ is the damping coefficient. The damping coefficient is derived from the proportion of two subsequent maxima of the oscillation $c_i(t_i)$ and $c_{i+1}(t_{i+1})$:

$$\delta = \frac{\ln\left(c_i/c_{i+1}\right)}{(t_{i+1} - t_i)}. \tag{26}$$

Usually the damping of the Chandler oscillation is expressed in terms of a quality factor Q. The reciprocal value Q^{-1} represents the specific dissipation, i.e. the loss of energy at the Chandler frequency (Munk and MacDonald 1960). The specific dissipation is related to the damping coefficient:

$$Q^{-1} = \frac{\delta\,(t_{i+1} - t_i)}{\pi}. \tag{27}$$

The numerical value of the quality factor that corresponds to the curve displayed in Fig. 6.6 ($k_2 = 0.35 + 0.0036i$) is $Q = 82$. In Table 6.1 values of period and quality factor of the Chandler oscillation from various studies are provided. They were computed from geodetic observations and models using different methods. Especially the quality factor is characterised by a high level of uncertainty.

If effects of gravitational and other geophysical processes are superposed, i.e. if torques, relative angular momenta and further deviations of the tensor of inertia are regarded in the Liouville equation, an interaction between forced and free oscillation occurs due to rotational deformations. While the impacts on the Chandler frequency are negligible (Okubo 1982; Jochmann 2003), the Chandler amplitude is strongly affected by the excitations (see Sect. 6.2).

Table 6.1 Periods and quality factors Q (with 90% confidence interval) of the Chandler oscillation from different studies

Chandler period	Q	[Interval]	Source
434.0 ± 2.5 days	100	[50, 400]	Wilson and Haubrich (1976)
431.7 days	24		Lenhardt and Groten (1985)
433.3 ± 3.6 days	179	[47, >1,000]	Wilson and Vicente (1990)
439.5 ± 1.2 days	72	[30, 500]	Kuehne et al. (1996)
433.7 ± 1.8 days	49	[35, 100]	Furuya and Chao (1996)
413 – 439 days			Schuh et al. (2001)
434.1 days	69		Seitz and Kutterer (2005)

6.4.2.2 Deformations Due to Surface Loads

Various processes in the subsystems of the Earth, such as the motion of atmospheric high-pressure and low-pressure systems, ocean bottom pressure changes or hydrologic variations over the continents due to flooding and snow, exert time-variable surface mass loads on the solid Earth. In this way they cause deformations of the Earth's body which are up to few centimetres in the vertical and several millimetres in the horizontal (Rabbel and Zschau 1985; Sun et al. 1995). The change of the surface geometry entails the redistribution of mass elements within the solid Earth which has a significant effect on both the Earth's gravity potential and its rotation. Consequently expedient information about atmosphere loading (van Dam and Herring 1994), non-tidal ocean loading (van Dam et al. 1997) and continental water storage variations (van Dam et al. 2001; Schuh et al. 2004) is required for an advanced interpretation and analysis of space geodetic observations (Rabbel and Schuh 1986; Manabe et al. 1991; Haas et al. 1997; Boehm et al. 2009). The surface forces exerted by time-variable mass distributions are in contrast to gravitationally induced body forces. While the latter cause large-scale and very regular deformations of the Earth that are well predictable, the effects of surface mass loads are mostly restricted to a few 100 km. Since they are irregular, they are hardly predictable (van Dam et al. 1997).

Vertical surface deformations of the solid Earth are usually computed following the theory of Farrell (1972). Pressure variations $p(\lambda, \varphi)$ (units of [Pa]) are related to time-variable surface mass loads $q(\lambda, \varphi)$ (units of [kg/m^2]) by

$$q(\lambda, \varphi) = \frac{p(\lambda, \varphi)}{g} \,, \tag{28}$$

where g is the gravitational acceleration. The radial displacement $d_{\mathrm{r}}(P)$ of the Earth at a position $P(\varphi_P, \lambda_P)$ caused by surface mass loads q_Q at locations $Q(\varphi_Q, \lambda_Q)$ on the Earth's surface area σ_Q is estimated by (Moritz and Mueller 1987)

$$d_{\mathrm{r}}(P) = \frac{a^3}{M} \iint\limits_{\sigma_Q} q_Q \sum_{n=0}^{\infty} h'_n P_n(\cos \psi_{PQ}) \, \mathrm{d}\sigma_Q \,. \tag{29}$$

In this equation h'_n denotes the degree n load Love number. The spherical distance between P and the location $Q(\varphi_Q, \lambda_Q)$ of an individual (point-)mass load is given by ψ_{PQ} which is the argument of the degree n Legendre Polynomial $P_n(\cos \psi_{PQ})$. More compact (29) can be written as

$$d_{\mathrm{r}}\,(P) = a^2 \iint\limits_{\sigma_Q} q_Q G\left(\psi_{PQ}\right) \, \mathrm{d}\sigma_Q \,, \tag{30}$$

where the abbreviation

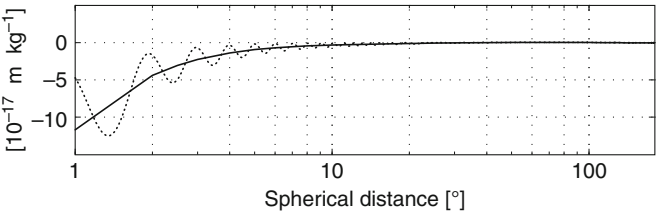


Fig. 6.7 Green's function $G\left(\psi_{PQ}\right)$ for the Earth model PREM. Dotted: truncation of the spherical harmonic expansion at degree $n = 350$, solid: infinite expansion

$$G\left(\psi_{PQ}\right) = \frac{a}{M} \sum_{n=0}^{\infty} h'_n P_n\left(\cos \psi_{PQ}\right) \tag{31}$$

is the Green's function for the vertical displacement (Farrell 1972). Function $G(\psi_{PQ})$ acts as a weighting operator which relates an individual surface mass load to the associated deformation of the solid Earth according to the spherical distance. Figure 6.7 shows the Green's function for continental crust computed from load Love numbers based on the previously mentioned Earth model PREM (Dziewonski and Anderson 1981; Scherneck 1990). The strong variability of the dotted curve truncated at $n = 350$ reflects the truncation error.

Figure 6.8 shows the time-variable deformations of the solid Earth for a period of two weeks in February 1994 as caused by atmosphere loading, non-tidal ocean loading and water storage variations over the continents (Seitz 2004). For the atmosphere and the ocean fields of surface mass loads $q_Q(\varphi_Q, \lambda_Q)$ were computed from a consistent combination of atmosphere surface pressure variations from reanalysis at the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) (Kalnay et al. 1996) and ocean bottom pressure variations from the constrained version kf049f of the global ocean circulation model ECCO (Fukumori 2002). Outputs of both models are provided in daily intervals; spatial resolutions are 2.5° × 2.5° for NCEP/NCAR (globally) and 1° × 1° for ECCO (between 70° N/S; densification of the grid around the equator). Since atmosphere pressure forcing is not taken into account by ECCO, an inverse barometric correction is applied to the NCEP/NCAR fields, i.e. air pressure is set to zero over the ocean. Variations of continental hydrology are taken from the land dynamics model (LaD; version *Euphrates*) (Milly and Shmakin 2002). LaD data comprehends monthly values of global water and groundwater storage as well as snow loads per 1° × 1° grid cell. While the deformations over the continents are up to 2 cm, the influence of ocean bottom pressure variations on the surface geometry of the Earth is marginal.

In order to assess the effect of the deformations on Earth rotation, the vertical surface displacements have to be transformed into variations of the tensor of inertia $\Delta\underline{\mathbf{I}}(t)$. Since this two-step procedure, i.e. the computation of global load deformations and the subsequent transformation of the deformations into deviations

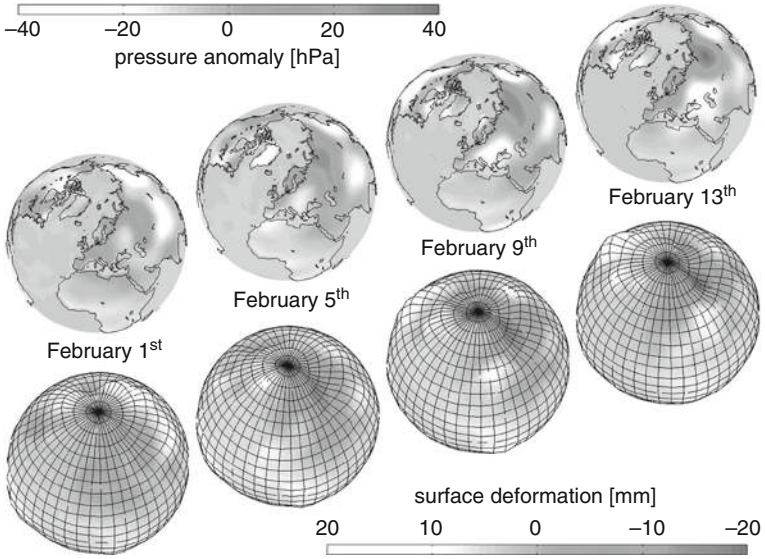


Fig. 6.8 Deformations of the solid Earth due to anomalies of atmospheric surface pressure, ocean bottom pressure, and continental water storage (Seitz 2004)

of the tensor of inertia (Dill 2002; Seitz 2004), is laborious and time-consuming, the indirect effect of mass redistributions on Earth rotation is commonly computed from changes of the geopotential associated with the mass load and the surface deformation.

The variation ΔU_{def} of the Earth's gravity potential U due to a surface deformation is proportional to the perturbing potential u of the surface mass load. The proportionality factor is the potential Love number k' (Moritz and Mueller 1987):

$$\Delta U_{\mathrm{def}} = k' u \; . \tag{32}$$

In general, the gravity potential u of a loading (point-)-mass m equals (Heiskanen and Moritz 1967)

$$u = \frac{Gm}{l} = \frac{Gm}{a} \sum_{n=0}^{\infty} \left(\frac{a - d_{\mathrm{r}}}{a} \right)^n P_n(\cos \psi_{PQ}) \; , \tag{33}$$

or, since the vertical deformation d_{r} is small compared to the Earth radius a:

$$u = \frac{Gm}{a} \sum_{n=0}^{\infty} P_n(\cos \psi_{PQ}) \; . \tag{34}$$

According to (32), the related change of the geopotential due to the induced surface deformation is

$$\delta U_{\mathrm{def}} = \frac{Gm}{a} \sum_{n=0}^{\infty} k'_n P_n(\cos \psi_{PQ}) \,. \tag{35}$$

If the point mass m is substituted by the (global) surface mass load q_Q, this equation turns into

$$\delta U_{\mathrm{def}} = Ga \iint_{\sigma_Q} q_Q \sum_{n=0}^{\infty} k'_n P_n(\cos \psi_{PQ}) \, \mathrm{d}\sigma_Q \,, \tag{36}$$

and the gravity potential u of the loading mass is

$$u = Ga \iint_{\sigma_Q} q_Q \sum_{n=0}^{\infty} P_n(\cos \psi_{PQ}) \, \mathrm{d}\sigma_Q \,. \tag{37}$$

Following the derivation given by Moritz and Mueller (1987), (37) can be written as

$$u = Ga \sum_{n=0}^{\infty} \frac{4\pi}{2n+1} q_{Qn} \,, \tag{38}$$

where q_{Qn} is the Laplace surface harmonic of degree n of the function q_Q, i.e.

$$q_Q = \sum_{n=0}^{\infty} q_{Qn} \,. \tag{39}$$

If accordingly the Laplace surface harmonic u_n of function u is introduced, the gravity potential of degree n of the surface mass load can be written as

$$u_n = Ga \frac{4\pi}{2n+1} q_{Qn} \,. \tag{40}$$

Since the variations of the tensor of inertia $\Delta \underline{\mathbf{I}}(t)$ are solely related to potential variations of degree 2 (Rochester and Smylie 1974; Chen et al. 2005; see also (22)), it is sufficient to evaluate (40) for $n = 2$. The temporal variation of $\delta U_2(t)$ due to the surface deformation is

$$\delta U_2(t) = k'_2 Ga \frac{4\pi}{5} q_{Q_2} \tag{41}$$

with $k'_2 = -0.308$ (Dong et al. 1996). The relation between the spherical harmonic coefficients of degree 2 and the elements of the Earth tensor of inertia is linear (Chen et al. 2005). Therefore tensor variations due to the indirect effect $\Delta \underline{\mathbf{I}}_{\mathrm{def}}(t)$ are computed simply by multiplying the direct tensor variations by the load Love number k'_2:

$$\Delta\underline{\mathbf{I}}_{\mathrm{def}}(t) = k'_2 \cdot \Delta\widehat{\underline{\mathbf{I}}}(t)\,, \tag{42}$$

where $\Delta\widehat{\underline{\mathbf{I}}}(t)$ denotes the direct tensor variations due to the mass redistributions within atmosphere, ocean and other subsystems that are causative for the Earth's surface deformations.

Given the above, the total (direct and indirect) effect of mass redistributions on the Earth's tensor of inertia is (Barnes et al. 1983)

$$\Delta\underline{\mathbf{I}}(t) = (1 + k'_2) \cdot \Delta\widehat{\underline{\mathbf{I}}}(t)\,. \tag{43}$$

Thus, the direct effect is attenuated by about 30% due to the deformation of the solid Earth. Note that this is only valid for mass redistributions that actually load the Earth's surface. For processes that are not accompanied by surface deformations (e.g. mass redistributions in the mantle) the load Love number k'_2 must be set to zero in this equation (Gross 2007).

6.4.3 Solution of the Euler–Liouville Equation

In order to compute variations of Earth rotation from angular momentum changes and torques, the Liouville equation (12) has to be solved for the unknown quantities $m_i(t)$ of the Earth rotation vector $\boldsymbol{\omega}(t)$. In general two different approaches, an analytical and a numerical approach, are applicable for the solution of the coupled system of the three first-order differential equations. Both methods will be discussed in the following.

Less compact the Liouville equation (12) can be written as

$$\dot{\underline{\mathbf{I}}}\,\boldsymbol{\omega} \;+\; \underline{\mathbf{I}}\,\dot{\boldsymbol{\omega}} \;+\; \dot{\mathbf{h}} \;+\; \boldsymbol{\omega} \times \underline{\mathbf{I}}\,\boldsymbol{\omega} \;+\; \boldsymbol{\omega} \times \mathbf{h} \;=\; \mathbf{L}\,, \tag{44}$$

where the dot denotes the derivative with respect to time. The individual terms read explicitly

$$\dot{\underline{\mathbf{I}}}\,\boldsymbol{\omega} \;=\; \begin{pmatrix} \dot{c}_{11} & \dot{c}_{12} & \dot{c}_{13} \\ \dot{c}_{12} & \dot{c}_{22} & \dot{c}_{23} \\ \dot{c}_{13} & \dot{c}_{23} & \dot{c}_{33} \end{pmatrix} \cdot \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix}\,, \tag{45}$$

$$\underline{\mathbf{I}}\,\dot{\boldsymbol{\omega}} \;=\; \begin{pmatrix} A + c_{11} & c_{12} & c_{13} \\ c_{12} & B + c_{22} & c_{23} \\ c_{13} & c_{23} & C + c_{33} \end{pmatrix} \cdot \Omega \begin{pmatrix} \dot{m}_1 \\ \dot{m}_2 \\ \dot{m}_3 \end{pmatrix}\,, \tag{46}$$

$$\dot{\mathbf{h}} \;=\; \begin{pmatrix} \dot{h}_1 \\ \dot{h}_2 \\ \dot{h}_3 \end{pmatrix}\,, \tag{47}$$

$$\boldsymbol{\omega} \times \underline{\mathbf{I}}\,\boldsymbol{\omega} \;=\; \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix} \times \begin{pmatrix} A + c_{11} & c_{12} & c_{13} \\ c_{12} & B + c_{22} & c_{23} \\ c_{13} & c_{23} & C + c_{33} \end{pmatrix} \cdot \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix}\,, \tag{48}$$

$$\boldsymbol{\omega} \times \mathbf{h} \;=\; \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix} \times \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} . \tag{49}$$

The traditionally applied analytical approach has been described and discussed in various publications (e.g. Munk and MacDonald 1960; Lambeck 1980; Wahr 1982; Barnes et al. 1983; Moritz and Mueller 1987; Gross 2007). Therefore only its basic principle shall be sketched in the following. In the numerical ansatz, the non-linear equation system is solved directly via numerical integration.

6.4.3.1 Linear Analytical Approach

In order to allow for a closed solution of the coupled system of differential equations (44), the following simplifications are commonly introduced (Lambeck 1980):

- With adequate accuracy, the Earth can be viewed as a biaxial, i.e. rotationally symmetric, body, so that the principal components A and B can be substituted by their average value $A' = (A + B)/2$. (Indeed the quotient of the difference between A and B and the absolute value of either of them amounts to only 2.2×10^{-5}; see Gross 2007.)
- Terms that contain products of the small quantities $m_i(t), c_{ij}(t)$ and $h_i(t)$ or their derivatives with respect to time are negligible (linearisation).

With these assumptions, the expansion of the expressions (45), (46), (48) and (49) results in

$$\underline{\dot{\mathbf{I}}}\,\boldsymbol{\omega} \;=\; \Omega \begin{pmatrix} \dot{c}_{13} \\ \dot{c}_{23} \\ \dot{c}_{33} \end{pmatrix} , \tag{50}$$

$$\underline{\mathbf{I}}\,\dot{\boldsymbol{\omega}} \;=\; \Omega \begin{pmatrix} A' \dot{m}_1 \\ A' \dot{m}_2 \\ C \dot{m}_3 \end{pmatrix} , \tag{51}$$

$$\boldsymbol{\omega} \times \underline{\mathbf{I}}\,\boldsymbol{\omega} \;=\; \Omega^2 \begin{pmatrix} m_2(C - A') - c_{23} \\ -m_1(C - A') + c_{13} \\ 0 \end{pmatrix} , \tag{52}$$

$$\boldsymbol{\omega} \times \mathbf{h} \;=\; \Omega \begin{pmatrix} -h_2 \\ h_1 \\ 0 \end{pmatrix} . \tag{53}$$

Insertion into the Liouville equation (44) yields

$$\begin{aligned}&\Omega\begin{pmatrix}\dot{c}_{13}\\ \dot{c}_{23}\\ \dot{c}_{33}\end{pmatrix}+\Omega\begin{pmatrix}A'\,\dot{m}_1\\ A'\,\dot{m}_2\\ C\,\dot{m}_3\end{pmatrix}+\begin{pmatrix}\dot{h}_1\\ \dot{h}_2\\ \dot{h}_3\end{pmatrix}\\ &+\Omega^2\begin{pmatrix}m_2(C-A')-c_{23}\\ -m_1(C-A')+c_{13}\\ 0\end{pmatrix}+\Omega\begin{pmatrix}-h_2\\ h_1\\ 0\end{pmatrix}=\begin{pmatrix}L_1\\ L_2\\ 0\end{pmatrix},\end{aligned}\tag{54}$$

or component-by-component

$$\dot{m}_1\cdot\frac{A'}{\Omega(C-A')}+m_2=\frac{1}{\Omega^2(C-A')}\cdot\left[L_1+\Omega^2c_{23}-\Omega\dot{c}_{13}+\Omega h_2-\dot{h}_1\right]\;=:\;\Psi_2\,,\tag{55}$$

$$\dot{m}_2\cdot\frac{A'}{\Omega(C-A')}-m_1\;=\frac{1}{\Omega^2(C-A')}\cdot\left[L_2-\Omega^2c_{13}-\Omega\dot{c}_{23}-\Omega h_1-\dot{h}_2\right]\;=:\;-\Psi_1\,,$$

$$\dot{m}_3=\frac{1}{\Omega C}\cdot\left[-\Omega\dot{c}_{33}-\dot{h}_3\right]\;=:\;\dot{\Psi}_3\,.\tag{57}$$

The terms containing the time-variable equatorial components of the external torques $L_1(t)$ and $L_2(t)$ as well as the quantities $c_{ij}(t), h_i(t)$ or their derivatives with respect to time are referred to as *excitation functions* $\Psi_i\,(i=1,2,3)$ (Munk and MacDonald 1960). Note that $L_3=0$ due to $A=B$; see (18). Variations of the tensor elements $c_{13}(t), c_{23}(t)$ and $c_{33}(t)$ describe the sum of all direct effects of mass redistributions in the various system components and the effects of solid Earth deformations due to tides, polar motion and surface loads. For the principal moments of inertia A' and C numerical values have to be introduced that account for the effect of core–mantle decoupling as discussed in Sect. 6.4.2.1.

Due to the linearisation, only three of the six components of the tensor of inertia appear in the excitation functions. Deviation moments $c_{11}(t), c_{22}(t)$ and $c_{12}(t)$ are neglected in the analytical approach. The axial component $m_3(t)$ of the Earth rotation vector $\boldsymbol{\omega}(t)$ is decoupled from the horizontal components. With adequate accuracy $\Delta\mathrm{LOD}(t)$ can be calculated independently from polar motion (see (15)). For the computation of polar motion, the first two differential equations are transformed into a complex equation (Lambeck 1980). Defining $m(t)=m_1(t)+im_2(t)$ and $\Psi(t)=\Psi_1(t)+i\Psi_2(t)$ yields

$$i\cdot\left(\dot{m}\frac{A'}{\Omega(C-A')}\right)+m=\Psi\,,\tag{58}$$

where $i=\sqrt{-1}$. For a rigid body and in the absence of external torques Ψ equals zero. Then the solution of (58) is

$$m=m^0e^{i\sigma t}\,,\tag{59}$$

with the abbreviation $\sigma = \Omega\frac{(C-A')}{A'}$ and the complex coordinate $m^0 = m(t_0)$ as initial condition for the epoch t_0. The free polar motion of a rigid Earth would be a prograde and undamped oscillation with amplitude $|m^0|$ and frequency σ, which corresponds to a period of 304 days (Euler period).

The observed frequency of the free polar motion (Chandler frequency) differs from the Euler frequency due to the deformability of the Earth's body. As a consequence of mantle anelasticity, rotational deformations are accompanied by a loss of energy due to friction (see Sect. 6.4.2.1). In order to account for this effect, σ is substituted by the complex quantity σ_{CW} (Lambeck 1980):

$$\sigma_{\mathrm{CW}} = \sigma_0 \left(1 + \frac{i}{2Q}\right). \tag{60}$$

Here σ_0 is the Chandler frequency and Q denotes the quality factor that describes the damping of the Chandler amplitude due to dissipation. Both quantities are explicitly predetermined in the analytical approach. Therefore the result is directly dependent on the choice of the numerical values of σ_0 and Q (Wilson and Haubrich 1976). Due to the assumption of rotational symmetry ($A = B$), the resulting free polar motion is circular. For a deformable Earth the solution of the Liouville equation follows from the convolution

$$m = e^{i\sigma_{\mathrm{CW}}t}\left[m^0 - i\sigma_{CW}\int_{-\infty}^{t}\Psi(\tau)e^{-i\sigma_{\mathrm{CW}}\tau}\,\mathrm{d}\tau\right]. \tag{61}$$

Alternative to the explicit computation of polar motion from the excitation functions $\Psi_1(t)$ and $\Psi_2(t)$, an indirect method is commonly applied, in which the so-called *geodetic excitation* is derived by an inverse convolution from the observed polar motion (Chao 1985; Brzezinski 1992). The comparison between the gravitational and geophysical processes and the geodetic observations is then performed on the basis of the excitation functions $\Psi_1(t)$ and $\Psi_2(t)$ without calculating $m_1(t)$ and $m_2(t)$. Since the Chandler oscillation is a priori reduced from the observations in the course of the computation of the geodetic excitation, the indirect method is just like the direct method dependent on the choice of the parameters σ_0 and Q.

6.4.3.2 Non-linear Numerical Approach

In the non-linear numerical approach the system of the three differential equations (44) is solved numerically. In contrast to the analytical approach, the Chandler oscillation is not explicitly predetermined with respect to its period σ_0 and quality factor Q. Instead, the free polar motion is modelled by considering the effect of the back-coupling mechanism of rotational deformations in $\Delta\underline{\mathbf{I}}(t)$. As described in Sect. 6.4.2.1, frequency and damping of the Chandler oscillation are closely related to the value of the complex Love-number k_2 in this case (Seitz and Kutterer 2005).

No set-up of linearised excitation functions is required when the system of differential equations is solved numerically. This is a major difference compared to the

analytical approach: the tensor of inertia $\underline{\mathbf{I}}(t)$, the vector of angular momenta $\mathbf{h}(t)$ and the vector of the external torques $\mathbf{L}(t)$ are directly introduced into the Liouville equation. Therefore the temporal variations of the deviation moments c_{11}, c_{22} and c_{12} are also considered in the tensor of inertia, and the previously neglected higher-order terms are contained in the solution. As a consequence, the first two differential equations are not decoupled from the third one, i.e. polar motion and ΔLOD(t) are solved simultaneously.

Furthermore the numerical approach allows for the introduction of a triaxial approximate tensor $\underline{\mathbf{I}}_0(A \neq B \neq C)$. The free polar motion of such an unsymmetrical gyro is no longer circular. But since the discrepancy between A and B is small, the numerical eccentricity of the ellipse described by the Earth rotation vector with respect to the Earth-fixed reference frame is only 0.10 if all gravitational and geophysical excitations are neglected. Its semi-minor axis is oriented towards the direction of the smallest principal moment of inertia A.

The Liouville equation is reformulated as a coupled system of three ordinary differential equations of the general form

$$\dot{\mathbf{m}}(t) = f\left(t, m_i(t)\right) , \tag{62}$$

$(i = 1, 2, 3)$ with

$$\dot{\mathbf{m}}(t) = \begin{pmatrix} \dot{m}_1(t) \\ \dot{m}_2(t) \\ \dot{m}_3(t) \end{pmatrix} , \tag{63}$$

and

$$f(t, m_i(t)) = \begin{pmatrix} f_1(t, m_1(t), m_2(t), m_3(t)) \\ f_2(t, m_1(t), m_2(t), m_3(t)) \\ f_3(t, m_1(t), m_2(t), m_3(t)) \end{pmatrix} . \tag{64}$$

Function $f(t, m_i(t))$ comprehends the tensor of inertia $\underline{\mathbf{I}}(t)$, relative angular momenta $\mathbf{h}(t)$ and torques $\mathbf{L}(t)$. Due to rotational deformations, the tensor of inertia includes deviations, which are dependent on m_1 and m_2 (cf. (21) and (22)). Consequently derivatives of these parameters with respect to time $\dot{m}_i$ appear in both terms $\underline{\mathbf{I}}\ \dot{\boldsymbol{\omega}}$ and $\dot{\underline{\mathbf{I}}}\ \boldsymbol{\omega}$ of the Liouville equation (44). In order to solve the Liouville equation for the unknown quantities m_i, their derivatives are assembled on the left-hand side of the differential equation. Therefore the tensor of inertia $\dot{\underline{\mathbf{I}}}$ is divided into two parts: one component, $\dot{\underline{\mathbf{I}}}_{\mathrm{R}}$, describes the effect of rotational deformations and depends on $\dot{m}_i$; the second component $\dot{\underline{\mathbf{I}}}_{\mathrm{G}}$ includes the geophysically induced mass redistributions in the fluid system components and in the solid Earth due to tidal deformations and load deformations. This second component $\dot{\underline{\mathbf{I}}}_{\mathrm{G}}$ is independent of $\dot{m}_i$. Consequently the Liouville equation can be written as

$$\dot{\underline{\mathbf{I}}}_{\mathrm{R}}\ \boldsymbol{\omega} + \underline{\mathbf{I}}\ \dot{\boldsymbol{\omega}} = \mathbf{L} - \dot{\underline{\mathbf{I}}}_{\mathrm{G}}\ \boldsymbol{\omega} - \dot{\mathbf{h}} - \boldsymbol{\omega} \times \mathbf{h} - \boldsymbol{\omega} \times \underline{\mathbf{I}}\ \boldsymbol{\omega} . \tag{65}$$

If the products of the m_i with the (very small) derivatives $\dot{m}_i$ are neglected, the term $\dot{\underline{\mathbf{I}}}_{\mathrm{R}}\,\boldsymbol{\omega}$ equals

$$\dot{\underline{\mathbf{I}}}_{\mathrm{R}}\,\boldsymbol{\omega} \;=\; \frac{\Omega^3 a^5}{3G}\begin{bmatrix} \Re(k_2)\cdot\dot{m}_1 + \Im(k_2)\cdot\dot{m}_2 \\ \Re(k_2)\cdot\dot{m}_2 - \Im(k_2)\cdot\dot{m}_1 \\ 0 \end{bmatrix} \tag{66}$$

$$\approx \frac{\Omega^3 a^5}{3G}\Re(k_2)\begin{bmatrix} \dot{m}_1 \\ \dot{m}_2 \\ 0 \end{bmatrix}. \tag{67}$$

Since the real part of k_2 is two orders of magnitude larger than its imaginary part, the products of $\Im(k_2)$ with $\dot{m}_i$ are neglected, too. The left-hand side of system (65) turns into

$$\dot{\underline{\mathbf{I}}}_{\mathrm{R}}\,\boldsymbol{\omega} \;+\; \underline{\mathbf{I}}\,\dot{\boldsymbol{\omega}} = \left(\frac{\Omega^3 a^5}{3G}\Re(k_2)\begin{bmatrix} 1&0&0\\0&1&0\\0&0&0 \end{bmatrix} + \Omega\,\underline{\mathbf{I}}\right)\begin{bmatrix} \dot{m}_1 \\ \dot{m}_2 \\ \dot{m}_3 \end{bmatrix} \tag{68}$$

$$=: \underline{\mathbf{F}}\begin{bmatrix} \dot{m}_1 \\ \dot{m}_2 \\ \dot{m}_3 \end{bmatrix}, \tag{69}$$

and the Liouville equation can be converted into

$$\dot{\mathbf{m}} = \underline{\mathbf{F}}^{-1}\left(\mathbf{L} \;-\; \dot{\underline{\mathbf{I}}}_{\mathrm{G}}\,\boldsymbol{\omega} \;-\; \dot{\mathbf{h}} \;-\; \boldsymbol{\omega}\times\mathbf{h} \;-\; \boldsymbol{\omega}\times\underline{\mathbf{I}}\,\boldsymbol{\omega}\right). \tag{70}$$

This coupled system of three first-order differential equations is solved as an initial value problem. Just as in the analytical approach, the tensor of inertia is composed of all contributions from direct mass effects in the components of the Earth system and from the deformations of the solid Earth induced by tides, polar motion and loading. The effect of core–mantle decoupling is regarded by the adoption of appropriate values of the principal moments of inertia A, B, C; see Sect. 6.4.2.1.

The introduction of initial values for the epoch t_0

$$\mathbf{m^0} = \begin{pmatrix} m_1^0 \\ m_2^0 \\ m_3^0 \end{pmatrix} \tag{71}$$

allows for a unique computation of special solutions for the unknown functions $m_i = m_i(t)$ by which the initial conditions $m_i(t_0) = m_i^0$ are fulfilled. Respective values m_i^0 are deduced from observations of polar motion and ΔLOD. For the correspondence between the geodetic observations and the quantities $m_i(t)$, see Sect. 6.5.

The efficiency of the numerical solution is naturally linked to the quality of both the applied initial conditions and the integrator. Using the previously mentioned dynamic Earth system model (see Sect. 6.4.2.1), Seitz and Kutterer (2005)

studied the effect of inaccurate initial values on the numerical solution. They concluded from 30 test runs that the variation of the initial values within an interval of $\pm 3\sigma_i$ (where σ_i is the standard deviation of the respective observation) is uncritical. However, if the applied initial values differ substantially from the geodetic observations (e.g. if $m_i^0 = 0$ is assumed), the results are contaminated over many decades. In this case the model time series are not interpretable before a steady state is reached, i.e. before the influence of the initial conditions falls below the level of the model accuracy. Studies on the reliability of the solution from an algorithmic point of view showed little dependence of the results on the choice of a specific solver: Seitz (2004) applied various one-step and multi-step solvers as well as an extrapolation method in the dynamic Earth system model and performed runs over more than two decades with identical initial values and forcing conditions. It was shown that the RMS differences between all runs were in the order of 1 mas for polar motion and 1–2 μs for ΔLOD. In contrast, the absolute accuracy of the model, i.e. the RMS difference between the simulation and the geodetic observations, was in the order of 30 mas for polar motion and 120 μs for ΔLOD.

More details on the non-linear approach, numerical model results from the dynamic Earth system model and comparisons between modelled and observed time series for polar motion and ΔLOD can be found in Seitz (2004) and Seitz and Schmidt (2005).

6.5 Relation Between Modelled and Observed Variations of Earth Rotation

The parameters $m_i(t)$ of the Earth rotation vector $\boldsymbol{\omega}(t)$ in the Liouville Equation (12) are related to the geodetically observed time series of polar motion $[x_\mathrm{p}(t), y_\mathrm{p}(t)]$, $\Delta\mathrm{UT}(t)$ and length-of-day variations $\Delta\mathrm{LOD}(t)$. The time-variable angular misalignment between the instantaneous rotation vector $\boldsymbol{\omega}(t)$ and the z-axis of the terrestrial reference frame in x- and y-direction is described by the two components $m_1(t)$ and $m_2(t)$, respectively (see (14)). On the other hand, the coordinates x_p and y_p published by the IERS represent the misalignment between the CIP and the IERS Reference Pole (i.e. the z-axis of the ITRF).

Both coordinate systems differ with respect to the direction of the positive y-axis: the terrestrial system, to which the rotation vector $\boldsymbol{\omega}(t)$ refers to, is a right-hand system. The system used by the IERS for the publication of the coordinates $x_\mathrm{p}(t)$ and $y_\mathrm{p}(t)$, however, is a left-hand system (y_p-axis directed towards 90°W; see Sect. 6.1). Therefore the coordinates $[p_1(t), p_2(t)]$ are defined, in order to describe the position of the CIP with respect to the IERS Reference Pole in a right-hand system:

$$\begin{aligned} p_1(t) &= x_\mathrm{p}(t)\,, \\ p_2(t) &= -y_\mathrm{p}(t)\,. \end{aligned} \tag{72}$$

The relation between $[m_1(t), m_2(t)]$ and $[p_1(t), p_2(t)]$ follows from the transformation between the true celestial equator system and the Earth-fixed system. According to Sect. 6.1, the rotation matrix $\underline{A}(t)$ that transforms between both systems accounts

for the daily spin around the axis of the CIP and polar motion:

$$\underline{A}(t) = \underline{W}(t)\, \underline{S}(t)\ . \tag{73}$$

The coordinates $\omega_1(t), \omega_2(t), \omega_3(t)$ of the Earth rotation vector $\boldsymbol{\omega}(t)$ in the terrestrial system are the elements of the skew-symmetric matrix $\underline{B}(t)$:

$$\underline{B}(t) = \begin{pmatrix} 0 & \omega_3(t) & -\omega_2(t) \\ -\omega_3(t) & 0 & \omega_1(t) \\ \omega_2(t) & -\omega_1(t) & 0 \end{pmatrix} = \dot{\underline{A}}(t)\underline{A}^{\mathrm{T}}(t)\ . \tag{74}$$

Since matrix $\underline{W}(t)$, and accordingly matrix $\underline{A}(t)$, do not include the previously mentioned Oppolzer motion, the Earth rotation vector $\boldsymbol{\omega}(t)$ resulting from (74) does not account for the subdaily retrograde deflection between the instantaneous rotation axis and the direction to the CIP. In order to include this effect, a vector that describes the departure between CIP and the instantaneous rotation pole (IRP) would have to be added to the product $\dot{\underline{A}}(t)\underline{A}^{\mathrm{T}}(t)$ (see below).

For time scales longer than 1 day, the comparison of the coefficients of matrix $\underline{B}(t)$ with the result of the product $\dot{\underline{A}}(t)\underline{A}^{\mathrm{T}}(t)$ leads to the relation between the elements of the Earth rotation vector $\boldsymbol{\omega}(t)$ and the coordinates $p_1(t)$ and $p_2(t)$. To the first order, this relation is (Brzezinski 1992; Gross 1992)

$$\begin{aligned} \omega_1(t) &= \Omega\, p_1(t) + \dot{p}_2(t)\,, \\ \omega_2(t) &= \Omega\, p_2(t) - \dot{p}_1(t)\ , \end{aligned} \tag{75}$$

or with $\omega_1(t) = \Omega\, m_1(t)$ and $\omega_2(t) = \Omega\, m_2(t)$

$$\begin{aligned} m_1(t) &= p_1(t) + \frac{1}{\Omega}\, \dot{p}_2(t)\,, \\ m_2(t) &= p_2(t) - \frac{1}{\Omega}\, \dot{p}_1(t)\ . \end{aligned} \tag{76}$$

The connection between modelled and observed polar motion is illustrated in Fig. 6.1 following Mendes Cerveira et al. (2009). The published polar motion values $p(t)$ refer to the position of the CIP in the ITRF, where $p(t) = x_p(t) - iy_p(t) = p_1(t) + ip_2(t)$. As stated above, the model values $m(t) = m_1(t) + im_2(t)$ represent polar motion of the IRP in the terrestrial frame that differs from the CIP by the effect of the Oppolzer motion (vector d in Fig. 6.1) (Capitaine 2002; Mendes Cerveira et al. 2009). For completeness, the axes of the space-fixed reference frame (GCRF) are also sketched. The coordinates $X(t)$ and $Y(t)$ of the CIP in the GCRF are derived from the precession–nutation (PN) model IAU 2000A and the published celestial pole offsets as described in Sect. 6.1.

The correspondence between the variation of the absolute value of the Earth rotation vector $\boldsymbol{\omega}(t)$ and the observed quantities $\Delta\mathrm{LOD}(t)$ and $\Delta\mathrm{UT}(t)$ results directly from the definition of $\Delta\mathrm{LOD}(t)$, meaning the time span of one full revolution of the Earth reduced by 86,400 s (see (5)):

$$\Delta \mathrm{LOD}\,(t) = \frac{2\,\pi\,\kappa}{|\boldsymbol{\omega}(t)|} - 86,400\,\mathrm{s}\,, \tag{77}$$

where

$$\kappa = \frac{\Omega}{2\,\pi} \cdot 86,400\,\mathrm{s} \quad = \quad \frac{86,400}{86,164}\,. \tag{78}$$

The introduction of the absolute value of $\boldsymbol{\omega}(t)$ (15) delivers (Schneider 1988)

$$\Delta \mathrm{LOD}\,t = \frac{2\,\pi\,\kappa}{\Omega(1 + m_3(t))} - 86,400\,\mathrm{s} \quad = \quad -m_3(t) \cdot 86,400\,\mathrm{s}\,, \tag{79}$$

and according to the relation between ΔLOD (t) and ΔUT (t) (6) (Gross 1992)

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta \mathrm{UT}\,(t) = m_3(t)\,. \tag{80}$$

Following the derivations in this section, physical model results and published values of Earth orientation parameters from space-geodetic observation techniques can be uniquely related to each other. In this way, physical models of Earth rotation can contribute significantly to the interpretation of the observations in terms of geophysical processes in the Earth system. Studies of the Earth's reaction on gravitational and other geophysical excitations, e.g. processes and interactions within and between atmosphere, hydrosphere and solid Earth can be performed using theoretical forward models. As described in Sect. 6.4, such models comprehend physical transfer functions that relate gravitational and geophysical model data and/or observations to time series of geodetic parameters of rotation, gravity field variations and changes of the surface geometry of the Earth. In forward models, observations of Earth rotation are used as a reference in order to examine the quality of geophysical data sets by balancing modelled angular momentum variations in the Earth's subsystems with the observed integral signal. Forward models have also been used for the prediction of geodetic parameters, e.g. in the context of global change, when climate predictions are introduced as forcing (Winkelnkemper et al. 2008).

Vice versa, observed time series of Earth orientation parameters can be used in order to support and improve theoretical models via inverse methods. In this way, the geodetic observations can contribute directly to an improved understanding of Earth system dynamics. Inverse methods have been developed for many years in geodesy. They are directed towards the gain of knowledge from precise geodetic observations about geophysical parameters (Marchenko and Schwintzer 2003), individual dynamic processes or interactions in the Earth system. While numerous recent studies deal with the assessment of the Earth's mass redistribution from an inversion of the time-variable gravity field from GRACE (Chao 2005; Ramillien et al. 2005), time series of geometric surface deformations from GPS (Wu et al. 2003) or combination of both (Wu et al. 2002; Kusche and Schrama 2005), there are hardly any approaches for the inversion of Earth rotation up to now. But due to the long observation records of polar motion and ΔLOD, due to the high accuracy

of the measurements and due to the large spectral range from hours to decades that is covered by the observations, the development of inverse Earth rotation models is a highly promising challenge for the future.

Acknowledgement The authors would like to express their gratitude to Urs Hugentobler from the Technische Universität München, Germany, and to Aleksander Brzezinski from the Polish Academy of Sciences, Warsaw, Poland, whose comments on the manuscript were very helpful and substantially improved the content of this chapter.

References

Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B. and Boucher, C. (2007) ITRF2005: a new release of the International Terrestrial Reference Frame based on time series of station positions and Earth Orientation Parameters. J. Geophys. Res., 112, 10.1029/2007JB004949

Aoki, S., Guinot, B., Kaplan, G.H., Kinoshita, H., McCarthy, D.D. and Seidelmann, P.K. (1982) The new definition of universal time. Astron. Astrophys., 105, 359–361

Aoki, S. and Kinoshita, H. (1983) Note on the relation between the equinox and Guinot's nonrotating origin. Celest. Mech. Dyn. Astr., 29, 335–360

Barnes, R.T.H., Hide, R.H., White, A.A. and Wilson, C.A. (1983) Atmospheric angular momentum fluctuations, length of day changes and polar motion. Proc. R. Soc. Lon., 387, 31–73

Beutler, G. (2005) Methods of Celestial Mechanics I: Physical, Mathematical and Numerical Principles. Springer, Berlin

BIPM (2007) Director's Report on the Activity and Management of the International Bureau of Weights and Measures, Bureau International des Poids et Mesures, sévres Cedex

Bizouard, C. and Gambis, D. (2009) The combined solution C04 for Earth Orientation Parameters consistent with International Terrestrial Reference Frame 2005. In: Drewes, H. (ed) Geodetic Reference Frames. IAG Symposia 134, Springer, Berlin

Boehm, J., Heinkelmann, R., Mendes Cerveira, J.P., Pany, A. and Schuh, H. (2009) Atmospheric loading corrections at the observation level in VLBI analysis, J. Geodesy, 83, 1107–1113

Brzezinski, A. (1992) Polar motion excitation by variations of the effective angular momentum function: considerations concerning deconvolution problem. manuscripta geodaetica, 17, 3–20

Brzezinski, A. (2001) Diurnal and subdiurnal terms of nutation: a simple theoretical model for a nonrigid Earth. In: Capitaine, N. (ed) Proceedings of the Journées Systémes de Référence Spatiotemporels 2000, Paris, pp. 243–251

Brzezinski, A. and Nastula, J. (2000) Oceanic excitation of the Chandler wobble. Adv. Space Res., 30(2), 195–200

Capitaine, N. (2002) Comparison of 'old' and 'new' concepts: the Celestial Intermediate Pole and Earth orientation parameters. In: Capitatine, N., Gambis, D., McCarthy, D.D., Petit, G., Ray, J., Richter, B., Rothacher, M., Standish, E.M. and Vondrak, J. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, IERS Technical Note 29, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 35–44

Capitaine, N. (2004) Oppolzer terms: a review, FGS Workshop on 'Ring Laser Gyroscopes and Earth Rotation', Wettzell

Capitaine, N. (2008) Definition and realization of the celestial intermediate reference system. Proc. IAU 2007, 3, 10.1017/S1743921308019583

Capitaine, N., Chapront, J., Lambert, S. and Wallace, P. (2002) Expressions for the coordinates of the CIP and the CEO Using IAU 2000 Precession-Nutation. In: Capitatine, N., Gambis, D., McCarthy, D.D., Petit, G., Ray, J., Richter, B., Rothacher, M., Standish, E.M. and Vondrak, J. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, IERS Technical Note 29, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 89–91

Chandler, S.C. (1891) On the variation of latitude I-IV. Astron. J., 11, 59–61, 65–70, 75–79, 83–86
Chandler, S.C. (1892) On the variation of latitude V-VII. Astron. J., 12, 17–22, 57–72, 97–101
Chao, B.F. (1985) On the excitation of the Earth's polar motion. Geophys. Res. Lett., 12(8), 526–529
Chao, B.F. (1989) Length-of-day variations caused by El Nino Southern Oscillation and quasibiennial oscillation. Science, 243, 923–925
Chao, B.F. (1994) The geoid and Earth rotation. In: Vanićek, P. and Christou, N.T. (eds) Geoid and Its Geophysical Interpretations. CRC Press, Boca Raton, pp. 285–298
Chao, B.F. (2005) On inversion for mass distribution from global (time-variable) gravity field. J. Geodyn., 39(3), 223–230
Chao, B.F. and Gross, R.S. (1987) Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes. Geophys. J. R. Astron. Soc., 91, 569–596
Chao, B.F. and Gross, R.S. (2005) Did the 26 December 2004 Sumatra, Indonesia, earthquake disrupt the Earth's rotation as the mass media have said? EOS Trans. Amer. Geophys. Union, 86, 1–2
Chen, J.L., Wilson, C.R. and Tapley, B.D. (2005) Interannual variability of low-degree gravitational change, 1980–2002. J. Geodesy, 78, 535–543
Dehant, V., Arias, F., Bizouard, C., Bretagnon, P., Brzezinski, A., Buffett, B., Capitaine, N., Defraigne, P., de Viron, O., Feissel, M., Fliegel, H., Forte, A., Gambis, D., Getino, J., Gross, R., Herring, T., Kinoshita, H., Klioner, S., Mathews, P., Mc-Carthy, D., Moisson, X., Petrov, S., Ponte, R., Roosbeek, F., Salstein, D., Schuh, H., Seidelmann, K., Soffel, M., Souchay, J., Vondrak, J., Wahr, J., Wallace, P., Weber, R., Williams, J., Yatskiv, Y., Zharov, V. and Zhu, S. (1999) Considerations concerning the non-rigid Earth nutation theory. Celest. Mech. Dyn. Astr., 72(4), 245–310
de Viron, O., Bizouard, C., Salstein, D. and Dehant, V. (1999) Atmospheric torque on the Earth and comparison with atmospheric angular momentum variations. J. Geophys. Res., 104, 4861–4875
de Viron, O. and Dehant, V. (2003a) Tests on the validity of atmospheric torques on Earth computed from atmospheric model outputs. J. Geophys. Res., 108, 10.1029/2001JB001196
de Viron, O. and Dehant, V. (2003b) Reliability of atmospheric torque for geodesy. In: Richter, B., Schwegmann, W. and Dick, W.R. (eds) Proceedings of the IERS Workshop on Combination Research and Global Geophysical Fluids, IERS Technical Note 30, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 125–126
de Viron, O., Koot, L. and Dehant, V. (2005) Polar motion models: the torque approach. In: Plag, H.-P., Chao, B.F., Gross, R. and van Dam, T. (eds) Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Variations, Cahiers du Centre Européen de Géodynamique et de Séismologie 24, Luxembourg
de Viron, O., Ponte, R.M. and Dehant, V. (2001) Indirect effect of the atmosphere through the oceans on the Earth nutation using the torque approach. J. Geophys. Res., 106, 8841–8851
Dill, R. (2002) Der Einfluss von Sekundäreffekten auf die Rotation der Erde, C 550, Deutsche Geodätische Kommission, München (in German)
Dong, D., Gross, R.S. and Dickey, J.O. (1996) Seasonal variations of the Earth's gravitational field: an analysis of atmospheric pressure, ocean tidal, and surface water excitation. Geophys. Res. Lett., 23(7), 725–728
Dziewonski, A.M. and Anderson, D.L. (1981) Preliminary Reference Earth model (PREM). Phys. Earth Planet. Int., 25, 297–356
Engels, J. and Grafarend, E.W. (1999) Zwei polare geodätische Bezugssysteme: Der Referenzrahmen der mittleren Oberflächenvortizität und der Tisserand-Referenzrahmen. In: Schneider, M. (ed) 3. DFG-Rundgespräch zum Thema Bezugssysteme. Mitteilungen des Bundesamts für Kartographie und Geodäsie, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main (in German), pp. 100–109
Euler, L. (1765) Du mouvement de rotation des corps solides autour d'un axe variable. Mémoires de l'académie des sciences de Berlin, 14, 154–193
Farrell, W. (1972) Deformation of the Earth by surface loads. Rev. Geophys. Space Phys., 10, 761–797

Feissel, M. and Mignard, F. (1998) The adoption of ICRS on 1 January 1998: meaning and consequences. Astron. Astrophys., 331, L33–L36
Fey, A.L., Ma, C., Arias, E.F., Charlot, P., Feissel-Vernier, M., Jacobs, A.M.G.C.S., Li, J. and MacMillan, D.S. (2004) The second extension of the International Celestial Reference Frame. Astron. J., 127, 3587–3608
Fricke, W., Schwan, H. and Lederle, T. (1988) Fifth Fundamental Catalogue, Part I., Techn. Ber., Veröff. Astron. Rechen Inst., Heidelberg
Fukumori, I. (2002) A partitioned Kalman filter and smoother. Mon. Weather Rev., 130, 1370–1383
Furuya, M. and Chao, B.F. (1996) Estimation of period and Q of the Chandler wobble. Geophys. J. Int., 127, 693–702
Furuya, M., Hamano, Y. and Naito, I. (1996) Quasi-periodic wind signal as a possible excitation of Chandler wobble. J. Geophys. Res., 101, 25537–25546
Gilbert, F. and Dziewonski, A.M. (1975) An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. Phil. Trans. R. Soc., A278, 187–269
Gipson, J.M. and Ma, C. (1998) Site displacement due to variation in Earth rotation. J. Geophys. Res., 103, 7337–7350
Gontier, A.M., Arias, E.F. and Barache, C. (2006) Maintenance of the ICRF using the most stable sources. In: Souchay, J. and Feissel-Vernier, M. (eds) The International Celestial Reference System and Frame, 7–19, IERS Technical Note 34, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main
Gross, R.S. (1986) The influence of earthquakes on the Chandler wobble during 1977–1983. Geophys. J. R. Astr. Soc., 85, 161–177
Gross, R.S. (1992) Correspondence between theory and observations of polar motion. Geophys. J. Int., 109, 162–170
Gross, R.S. (1993) The effect of ocean tides on the Earth's rotation as predicted by the results of an ocean tide model. Geophys. Res. Lett., 20(4), 293–296
Gross, R.S. (2000) The excitation of the Chandler wobble. Geophys. Res. Lett., 27(15), 2329–2332
Gross, R.S. (2007) Earth rotation variations – Long period. In: Herring, T.A. (ed) Physical Geodesy, Treatise on Geophysics, Vol. 3. Elsevier, Amsterdam
Guinot, B. (1979) Basic problems in the kinematics of the rotation of the earth. In: McCarthy, D. and Pilkington, J. (eds) Time and the Earth's Rotation. D. Reidel, Dordrecht, pp. 7–18
Guinot, B. (2002) Comparison of 'old' and 'new' concepts: Celestial Ephemeris Origin (CEO), Terrestrial Ephemeris Origin (TEO), Earth Rotation Angle (ERA). In: Capitatine, N., Gambis, D., McCarthy, D.D., Petit, G., Ray, J., Richter, B., Rothacher, M., Standish, E.M. and Vondrak, J. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, IERS Technical Note 29, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 45–50
Haas, R., Scherneck, H.-G. and Schuh, H. (1997) Atmospheric loading corrections in geodetic VLBI and determination of atmospheric loading coefficients. In: Pettersen, B.R. (ed) Proceedings of the 12th Working Meeting on European VLBI for Geodesy and Astronomy, Statens Kartverk, Hønefoss, pp. 111–121
Hameed, S. and Currie, R.G. (1989) Simulation of the 14-month Chandler wobble climate model. Geophys. Res. Lett., 16(3), 247–250
Heiskanen, W.A. and Moritz, H. (1967) Physical Geodesy. Freeman and Co., San Francisco
Hinderer, J., Legros, H., Gire, C. and Le Mouël, J.-L. (1987) Geomagnetic secular variation, core motions and implications for the Earth's wobbles. Phys. Earth Planet. Int., 49, 121–132
Holme, R. (1998) Electromagnetic core-mantle coupling – I. Explaining decadal changes in length of day. Geophys. J. Int., 132, 167–180
Höpfner, J. (2001) Interannual variations in length of day and atmospheric angular momentum with respect to ENSO cycles. Zeitschrift f. Vermessungswesen, 126(1), 39–49
IERS (2008) IERS Annual Report 2006. In: Dick, W.R. und Richter, B. (Hrsg.) Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main

Jackson, A. (1997) Time-dependency of geostrophic core surface motions. Phys. Earth Planet. Int., 103, 293–311

Jault, D., Gire, C. and Le Mouël, J.-L. (1988) Westward drift, core motions and exchanges of angular momentum between core and mantle. Nature, 333, 353–356

Jochmann, H. (2003) Period variations of the Chandler wobble. J. Geodesy, 77, 454–458

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Wollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. (1996) The NMC/NCAR 40-year reanalysis project. Bull. Am. Meteor. Soc., 77, 437–471

Kaplan, G.H. (2005) The IAU resolutions on astronomical reference systems, time scales, and Earth rotation models, Techn. Ber. Circular 179, United States Naval Observatory, Washington

Kosek, W., McCarthy, D.D. and Luzum, B. (2001) El Niño impact on polar motion prediction errors. Studia Geophysica et Geodaetica, 45, 347–361

Kuehne, J., Wilson, C.R. and Johnson, S. (1996) Estimates of the Chandler wobble frequency and Q. J. Geophys. Res., 101, 13573–13580

Kusche, J. and Schrama, E. (2005) Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data. J. Geophys. Res., 110, 10.1029/2004JB003556

Lambeck, K. (1980) The Earth's Variable Rotation: Geophysical Causes and Consequences. Cambridge University Press, Cambridge

Lenhardt, H. and Groten, E. (1985) Chandler wobble parameters from BIH and ILS data. Manuscripta Geodaetica, 10, 296–305

Liao, D.C. and Greiner-Mai, H. (1999) A new ΔLOD series in monthly intervals (1892.0–1997.0) and its comparison with other geophysical results. J. Geodesy, 73, 466–477

Lieske, J.H., Lederle, T., Fricke, W. and Morando, B. (1977) Expression for the precession quantities based upon the IAU (1976) system of astronomical constants. Astron. Astrophys., 58, 1–16

Ma, C., Arias, E.F., Eubanks, T.M., Fey, A.L., Gontier, A.-M., Jacobs, C.S., Sovers, O.J., Archinal, B.A. and Charlot, P. (1998) The International Celestial Reference Frame as realized by Very Long Baseline Interferometry. Astron. J., 116, 516–546

Manabe, S., Sato, T., Sakai, S. and Yokoyama, K. (1991) Atmospheric loading effects on VLBI observations. Proceedings of the AGU Chapman Conference on Geodetic VLBI, NOAA Technical Report NOS 137 NGS 49, Rockville, pp. 111–122

Marchenko, A.N. and Schwintzer, P. (2003) Estimation of the Earth's tensor of inertia from recent global gravity field solutions. J. Geodesy, 76, 495–509

Mathews, P.M., Buffet, B.A., Herring, T.A. and Shapiro, I.I. (1991) Forced nutations of the Earth: influences of inner core dynamics, part 2: numerical results and comparisons. J. Geophys. Res., 96, 8243–8257

Mathews, P.M., Herring, T.A. and Buffet, B.A. (2002) Modeling of nutation and precession: new nutation series for nonrigid Earth and insights into the Earth's interior. J. Geophys. Res., 107, 10.1029/2001JB000390

Mendes Cerveira, P.J., Boehm, J., Schuh, H., Klügel, T., Velikoseltsev, A., Schreiber, U. and Brzezinski, A. (2009) Earth rotation observed by Very Long Baseline Interferometry and ring laser. Pure Appl. Geophys., 166, 1499–1517

McCarthy, D.D. and Capitaine, N. (2002) Practical Consequences of Resolution B1.6 'IAU2000 Precession-Nutation Model', Resolution B1.7 'Definition of Celestial Intermediate Pole', Resolution B1.8 'Definition and Use of Celestial and Terrestrial Ephemeris Origin'. In: Capitatine, N., Gambis, D., McCarthy, D.D., Petit, G., Ray, J., Richter, B., Rothacher, M., Standish, E.M. and Vondrak, J. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, IERS Technical Note 29, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 9–17

McCarthy, D.D. and Petit, G. (eds) (2004) IERS Conventions 2003, IERS Technical Note 32, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main

McClure, P. (1973) Diurnal polar motion, Techn. Ber. GSFC Rep. X-529-73-259, Goddard Space Flight Cent., Greenbelt

Milly, P.C. and Shmakin, A.B. (2002) Global modeling of land water and energy balances. Part I: the land dynamics (LaD) model. J. Hydrometeorol., 3(3), 283–299

Milne, G.A. and Mitrovica, J.X. (1998) Postglacial sea-level change on a rotating Earth. Geophys. J. Int., 133, 1–19

Molodensky, M.S. (1961) The theory of nutation and diurnal Earth tides. Comm. Obs. R. Belgique, 142, 25–56

Moritz, H. and Mueller, I.I. (1987) Earth rotation. Theory and Observation. Ungar, New York

Morrison, L.V. and Stephenson, F.R. (1998) The sands of time and tidal friction. In: Brosche, P., Dick, W.R., Schwarz, O. and Wielen, R. (eds) The Message of the Angles – Astrometry from 1798 to 1998, Thun-Verlag, Frankfurt am Main, pp. 100–113

Munk, W.H. and MacDonald, G.J.F. (1960) The Rotation of the Earth. A Geophysical Discussion. Cambridge University Press, Cambridge

Okubo, S. (1982) Is the Chandler period variable? Geophys. J. R. Astr. Soc., 71, 629–646

Ponsar, S., Dehant, V., Holme, R., Jault, D., Pais, A. and Hoolst, T.V. (2002) The core and fluctuations in the Earth's rotation. In: Dehant, V., Creager, K.C., Karato, S. and Zatman, S. (eds) Earth's Core: Dynamics, Structure, Rotation. Geodynamics Series 31, American Geophysical Union, Washington, pp. 251–261

Rabbel, W. and Schuh, H. (1986) The influence of atmospheric loading on VLBI experiments. J. Geophys., 59, 164–170

Rabbel, W. and Zschau, J. (1985) Static deformations and gravity changes at the Earth's surface due to atmospheric loading. J. Geophys., 56, 81–99

Ramillien, G., Frappart, F., Cazenave, A. and Güntner, A. (2005) Time variations of land water storage from an inversion of 2 years of GRACE geoids. Earth Planet. Sci. Lett., 235, 283–301

Ray, J.R. (1996) Measurements of length of day using the Global Positioning System. J. Geophys. Res., 101, 20141–20149

Richter, B. (1995) Die Parametrisierung der Erdorientierung, Zeitschrift f. Vermessungswesen, 102(3), 109–119 (in German)

Rochester, M.G. and Smylie, D.E. (1974) On changes in the trace of the Earth's inertia tensor. J. Geophys. Res., 79, 4948–4951

Rosen, R.D., Salstein, D.A., Eubanks, T.M., Dickey, J.O. and Steppe, J.A. (1984) An El Nino signal in atmospheric angular momentum and Earth rotation. Science, 225, 411–414

Rothacher, M. (2002) Future IERS products: Implementation of the IAU 2000 Resolutions. In: Capitatine, N., Gambis, D., McCarthy, D.D., Petit, G., Ray, J., Richter, B., Rothacher, M., Standish, E.M. and Vondrak, J. (eds) Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions, IERS Technical Note 29, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, pp. 77–84

Sasao, T., Okubo, S. and Saito, M. (1980) A simple theory on the dynamical effects of a stratified fluid core upon nutational motion of the Earth. In: Fedorov, E.P., Smith, M.L. and Bender, P.L. (eds) Nutation and the Earth's Rotation. IAU Symposia 78, D. Reidel, Kiev, pp. 165–183

Scherneck, H.G. (1990) Loading Green's functions for a continental shield with a Q-structure for the mantle and density constraints from the geoid. Bull. d'Inform. Marées Terr., 108, 7775–7792

Schneider, M. (1988) Satellitengeodäsie, BI Wissenschaftsverlag, Zürich (in German)

Schödlbauer, A. (2000) Geodätische Astronomie: Grundlagen und Konzepte, de Gruyter, Berlin (in German)

Schreiber, U., Velikoseltsev, A., Rothacher, M., Klügel, T., Stedman, G. and Wiltshire, D. (2004) Direct measurement of diurnal polar motion by ring laser gyroscopes. J. Geophys. Res., 109, 10.1029/2003JB002803

Schuh, H., Dill, R., Greiner-Mai, H., Kutterer, H., Müller, J., Nothnagel, A., Richter, B., Rothacher, M., Schreiber, U. and Soffe, M. (eds) (2003) Erdrotation und globale dynamische Prozesse,

Mitteilungen des Bundesamts für Kartographie und Geodäsie, Band 32, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main (in German)
Schuh, H., Estermann, G., Crétaux, J.-F., Bergé-Nguyen, M. and van Dam, T. (2004) Investigation of hydrological and atmospheric loading by space geodetic techniques. In: Hwang, C., Shum, C.K. and Li, J.C. (eds) Satellite Altimetry for Geodesy, Geophysics and Oceanography. IAG Symposia 126, Springer, Berlin, pp. 123–132
Schuh, H., Nagel, S. and Seitz, T. (2001) Linear drift and periodic variations observed in long time series of polar motion. J. Geodesy, 74, 701–710
Seidelmann, P.K. (ed) (1992) Explanatory Supplement to the Astronomical Almanac. University Science Books, Mill Valley
Seitz, F. (2004) Atmosphärische und ozeanische Einflüsse auf die Rotation der Erde – Numerische Untersuchungen mit einem dynamischen Erdsystemmodell, C 578, Deutsche Geodätische Kommission, München (in German)
Seitz, M. (2009) Kombination geodätischer Raumbeobachtungsverfahren zur Realisierung eines terrestrischen Referenzsystems, C 630, Deutsche Geodätische Kommission, München (in German)
Seitz, F. and Kutterer, H. (2005) Sensitivity analysis of the non-linear Liouville equation. In Sansò, F. (ed) A Window on the Future of Geodesy. IAG Symposia 128, Springer, Berlin, pp. 601–606
Seitz, F. and Schmidt, M. (2005) Atmospheric and oceanic contributions to Chandler wobble excitation determined by wavelet filtering. J. Geophys. Res., 110, 10.1029/2005JB003826
Seitz, F., Stuck, J. and Thomas, M. (2004) Consistent atmospheric and oceanic excitation of the Earth's free polar motion. Geophys. J. Int., 157, 25–35
Sidorenkov, N.S. (1992) Excitation mechanisms of Chandler polar motion. Astr. Zh., 69(4), 905–909
Smith, M.L. and Dahlen, F.A. (1981) The period and Q of the Chandler wobble. Geophys. J. R. Astr. Soc., 64, 223–281
Souchay, J., Loysel, B., Kinoshita, H. and Folgueira, M. (1999) Corrections and new developments in rigid Earth nutation theory. Final tables 'REN-200' including crossed-nutation and spin-orbit coupling effects. Astron. Astrophys. Suppl. Ser., 135, 111–131
Souriau, A. and Cazenave, A. (1985) Reevaluation of the seismic excitation of the Chandler wobble from recent data. Earth Plan. Sci. Lett., 75, 410–416
Standish, E.M. (1998) JPL planetary and lunar ephemerides DE405/LE405, Techn. Ber. IOM 312.F-98-048, JPL, Pasadena
Stuck, J. (2002) Die simulierte axiale atmosphärische Drehimpulsbilanz des ECHAM3-T21 GCM, Bonner Meteorologische Abhandlungen 56, Asgard-Verlag, Sankt Augustin (in German)
Sun, H.P., Ducarme, B. and Dehant, V. (1995) Effect of the atmospheric pressure on surface displacements. J. Geodesy, 70, 131–139
Tisserand, F. (1891) Traité de Méchanique Céleste, Vol. II. Gauthier-Villars, Paris (in French)
Torge, W. (2001) Geodesy. de Gruyter, Berlin
Trenberth, K.E. (1980) Atmospheric quasibiennial oscillations. Mon. Weather Rev., 108, 1370–1377
van Dam, T.M. and Herring, T.A. (1994) Detection of atmospheric pressure loading using very long baseline interferometry measurements. J. Geophys. Res., 99, 4505–4517
van Dam, T.M., Wahr, J., Chao, Y. and Leuliette, E. (1997) Predictions of crustal deformation and of geoid and sea level variability caused by oceanic and atmospheric loading. Geophys. J. Int., 99, 507–515
van Dam, T.M., Wahr, J., Milly, P.C.D., Shmakin, A.B., Blewitt, G., Lavalee, D. and Larson, K.M. (2001) Crustal displacements due to continental water loading. Geophys. Res. Lett., 28, 651–654
Vondrak, J., Ron, C., Pesek, I. and Cepek, A. (1995) New global solution of Earth orientation parameters from optical astrometry in 1900–1990. Astron. Astrophys., 297, 899–906
Vondrak, J., Weber, R. and Ron, C. (2005) Free core nutation: direct observations and resonance effects. Astron. Astrophys., 444, 297–303

Wahr, J.M. (1981) The forced nutations of an elliptical, rotating, elastic and oceanless Earth. Geophys. J. R. Astr. Soc., 64, 705–727
Wahr, J.M. (1982) The effects of the atmosphere and oceans on the Earth's wobble – I. Theory. Geophys. J. R. Astr. Soc., 70, 349–372
Wahr, J.M. (1983) The effects of the atmosphere and the oceans on the Earth's wobble and on the seasonal variations in the length of day – II. Results. Geophys. J. R. Astr. Soc., 74, 451–487
Wahr, J.M. (1985) Deformation induced by polar motion. J. Geophys. Res., 90, 9363–9368
Wilson, C.R. and Haubrich, R.A. (1976) Meteorological excitation of the Earth's wobble. Geophys. J. R. Astr. Soc., 46, 707–743
Wilson, C.R. and Vicente, R.O. (1990) Maximum likelihood estimates of polar motion parameters. In: McCarthy, D.D. and Carter, W.E. (eds) Variations in Earth Rotation. Geophysical Monograph Series 59, American Geophysical Union, Washington, pp. 151–155
Winkelnkemper, T., Seitz, F., Min, S. and Hense, A. (2008) Simulation of historic and future atmospheric angular momentum effects on length-of-day variations with GCMs. In: Sideris, M. (ed) Observing our Changing Earth. IAG Symposia 133, Springer, Berlin, pp. 447–454
Wu, X., Heflin, M., Ivins, E., Argus, D. and Webb, F. (2003) Large-scale global surface mass variations inferred from GPS measurements of load-induced deformation. Geophys. Res. Lett., 30(14), 10.1029/2003GL017546
Wu, X., Watkins, M., Ivins, E., Kwok, R., Wang, P. and Wahr, J. (2002) Toward global inverse solutions for current and past ice ass variations: contribution of secular satellite gravity and topography change measurements. J. Geophys. Res., 107, 10.1029/2001JB000543
Yoder, C.F., Williams, J.G. and Parke, M.E. (1981) Tidal variations of Earth rotation. J. Geophys. Res., 86, 881–891

Chapter 7
Equivalence of GPS Algorithms and Its Inference

Guochang Xu, Yunzhong Shen, Yuanxi Yang, Heping Sun, Qin Zhang, Jianfeng Guo and Ta-Kang Yeh

Contents

G. Xu (✉)
GFZ German Research Centre for Geosciences, Department 1: Geodesy and Remote Sensing, Telegrafenberg, 14473 Potsdam, Germany
e-mail: xu@gfz-potsdam.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_7,

7.1 Introduction

In the past decade of GPS research, advantages and disadvantages of the differential and un-differential GPS have been discussed in detail in many publications. First in 2002 it was algebraically pointed out that the un-differential and differential algorithms of GPS are equivalent (Xu 2002a). Various combined and uncombined methods of GPS algorithms were discussed in detail too to show the advantages of some combining ways. However, again the combined and uncombined algorithms were proved to be equivalent (Xu et al. 2006a, 2007). The equivalence principle can now be easily explained and accepted, because the used GPS data and model are the same and the used adjustment principle is also the same. The information contents are the same; therefore the results one may obtain should also be equivalent.

If one accepts that the different GPS algorithms of differencing or combining are equivalent, then several problems arising from the equivalence principle should be studied too. One of the most important consequences is that the un-differential GPS model is a rank-deficient one. An independent parameterisation method is then developed (Xu et al. 2006b, 2007). Based on the above three developments, an equivalence theorem can be summarised (Xu et al. 2006c, 2007).

One of the most important by-products of the equivalence principle is the diagonalisation algorithm (Xu 2003a, 2007). It is even important for classic sequential adjustment in case of parameter reduction. The separability of any observation equation and its normal equation is another inference of the equivalence principle (Xu 2007). The optimal criterion of ambiguity search is the third (Xu 2002b, 2007).

The above factors formed part of the very important theoretical developments on GPS theoretical research, which are described in Xu (2007) in different sections throughout the book. A summary of the equivalence principle and emphasis of the basic principle are still important. This is the purpose of this chapter.

This chapter consists of seven sections. The equivalence of the differencing and undifferenced GPS algorithms is outlined theoretically in the second section. The by-product of the theory is that a unified GPS data processing method is proposed which may include all undifferenced and differencing algorithms into a unique one and any algorithm can be realised by using a switch. Such equivalence is soft and the definition is given. The differencing includes single-differencing methods, double-differencing methods and triple-differencing methods, as well as user-defined differencing. Incidentally, the triple-difference equations are equivalent

to the others only for real-valued ambiguities, since the equivalence principle is not applicable to the integer parameters.

The equivalence properties of the un-combined and combining GPS data processing algorithms are verified theoretically in the third section. The solution vector and the covariance matrix of the solutions are identical regardless of the algorithms adopted. The correct choice of special algorithms will lead to a better and easier dealing of the data and solving of the problem. Such equivalence is an exact one.

An independent parameterisation method of GPS observation model is discussed in the fourth section. Two sets of evidence of the singularity of the undifferenced GPS observation models are given to show the need for the study on the parameterisation problem. Then a general method to model the bias parameters independently is proposed through analysis. A geometry-free illustration is given to show the problem of singularity of the bias parameterisation. A correlation analysis in the case of a phase–code combination is also discussed to emphasise that the singularity will not change much because of the lower weight of the code.

The equivalence of the mixtures of the combining and differencing methods is discussed in the fifth section. A derivation is given to show that the traditional combinations are non-equivalent algorithms due to the inexact bias parameterisation. Equivalence property of the GPS data processing algorithms is summarised based on the equivalence of the uncombined and combining, undifferenced and differencing algorithms, as well as their mixtures.

Inferences of the equivalence principle are discussed in the sixth section, including the diagonalisation algorithm, the separability of any observation equation and its normal equation, the optimal criterion of ambiguity search. A short summary is given at the end.

7.2 Equivalence of Undifferenced and Differencing Algorithms

A unified GPS data processing method based on equivalently eliminated equations is described and the equivalence between undifferenced and differencing algorithms is proved in this section. The theoretical background of the method is given. By selecting the eliminated unknown vector as a zero vector, a vector of satellite clock error, a vector of all clock error, a vector of clock and ambiguity parameters or a vector of user-defined unknowns, the selectively eliminated equivalent observation equations can be formed. The equations are equivalent to the zero-differencing, single-differencing, double-differencing, triple-differencing or user-defined differencing equations. The advantage of such a method is that the different GPS data processing methods are unified to a unique one, wheras the observational vector remains the original one and the weight matrix keeps the un-correlated diagonal form. In other words, by using this equivalent method, one may selectively reduce the unknown number; however, one does not have to deal with the complicated correlation problem. Several special cases of single-, double- and triple-difference

are discussed in detail to illustrate the theory. The reference-related parameters are dealt with using the a priori datum method.

7.2.1 Introduction

In GPS data processing practice, the commonly used methods are so-called zero-difference (un-differential) methods, single-difference methods, double-difference methods and triple-difference methods (Hofmann-Wellenhof et al. 2001; Leick 2004; Remondi 1984; Strang and Borre 1997; Parkinson and Spilker 1996; Cannon et al. 1997; Petovell0 2006; Sien-Chong and William 1993; Sjoeberg 1999). It is well known that the observation equations of the differencing methods can be obtained by carrying out a related linear transformation to the original equations. As soon as the weight matrix is similarly transformed according to the law of covariance propagation, all methods are equivalent, theoretically. A theoretical proof of the equivalence between the un-differential and differential methods can be found in Schaffrin and Grafarend (1986). A comparison of the advantages and disadvantages of the un-differential and differential methods can be found, e.g., in de Jong (1998). The advantage of the differential methods is that the unknown parameters are fewer so that the whole problem to be solved becomes smaller. The disadvantage of the differential methods is that there is a correlation problem, which appears in cases of multiple baselines of single-difference methods and all double-difference as well as triple-difference methods. The correlation problem is often complicated and not easy to deal with exactly (compared with the un-correlated problem). The advantages and disadvantages reach a balance. If one wants to deal with a reduced problem (cancellation of many unknowns), then one has to deal with the correlation problem. As an alternative, we use the equivalent observation equation approach to unify the un-differential and differential methods, while keeping all the advantages of the un-differential and differential methods.

In the next sections, the theoretical basis of the equivalently eliminated equations will be given based on the derivation of Zhou (1985). Several detailed cases are then discussed to illustrate the theory. The reference-related parameters are dealt with using the a priori datum method. A summary of the selectively eliminated equivalent GPS data processing method is outlined at the end.

7.2.2 Formation of Equivalent Observation Equations

For the convenience of later discussion, the method to form an equivalently eliminated equation system is outlined here (cf. Sect. 7.6 of Xu 2007). In practice, sometimes only one group of unknowns is of interest; it is better to eliminate the other group of unknowns (called nuisance parameters), for example because of their size. In this case, using the so-called equivalently eliminated observation equation system could be very beneficial (Zhou 1985). The nuisance parameters

can be eliminated directly from the observation equations instead of from the normal equations.

The linearised observation equation system can be represented using the matrix

$$V = L - (A \quad B)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad P\,, \tag{1}$$

where L is an observation vector of dimension n, A and B are coefficient matrices of dimensions $n\times(s–r)$ and $n\times r$, X_1 and X_2 are unknown vectors of dimensions $s–r$ and r, V is residual error, s is the total number of unknowns and P is the weight matrix of dimension $n\times n$.

The related least squares normal equation can then be formed as

$$\begin{pmatrix} A & B \end{pmatrix}^{\mathrm{T}} P \begin{pmatrix} A & B \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A & B \end{pmatrix}^{\mathrm{T}} PL \tag{2}$$

or

$$M_{11}X_1 + M_{12}X_2 = B_1 \quad \text{and} \tag{3}$$

$$M_{21}X_1 + M_{22}X_2 = B_2, \tag{4}$$

where

$$B_1 = A^{\mathrm{T}}PL\,, \qquad B_2 = B^{\mathrm{T}}PL \quad \text{and}$$

$$\begin{pmatrix} A^{\mathrm{T}}PA & A^{\mathrm{T}}PB \\ B^{\mathrm{T}}PA & B^{\mathrm{T}}PB \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \tag{5}$$

After eliminating the unknown vector X_1, the eliminated equivalent normal equation system is then

$$M_2X_2 = R_2, \tag{6}$$

where

$$M_2 = -M_{21}M_{11}^{-1}M_{12} + M_{22} = B^{\mathrm{T}}PB - B^{\mathrm{T}}PAM_{11}^{-1}A^{\mathrm{T}}PB \quad \text{and} \tag{7}$$

$$R_2 = B_2 - M_{21}M_{11}^{-1}B_1\,. \tag{8}$$

The related equivalent observation equation of (6) is then (cf. Sect. 7.6 of Xu 2007; Zhou 1985)

$$U_2 = L - (E - J)BX_2, \quad P, \tag{9}$$

where

$$J = AM_{11}^{-1}A^{\mathrm{T}}P. \tag{10}$$

E is an identity matrix of size n, L and P are the original observation vector and weight matrix, respectively, and U_2 is the residual vector, which has the same property as V in (1). The advantage of using (9) is that the unknown vector X_1 has been eliminated; however, L vector and P matrix remain the same as the original.

Similarly, the X_2-eliminated equivalent equation system is

$$U_1 = L - (E - K)AX_1 \quad \text{and} \quad P, \tag{11}$$

where

$$K = BM_{22}^{-1}B^{\mathrm{T}}P\,, \qquad M_{22} = B^{\mathrm{T}}PB$$

and U_1 is the residual vector (which has the same property as V).

We have separated the observation equation (1) into two equations, (9) and (11); each equation contains only one of the unknown vectors. Each unknown vector can be solved independently and separately. Equations (9) and (11) are called equivalent observation equations of (1).

The equivalence property of (1) and (9) is valid under three implicit assumptions. The first is that the identical observation vector is used. The second is that the parameterisation of X_2 is identical. The third is that the X_1 is able to be eliminated. Otherwise, the equivalence does not hold.

7.2.3 Equivalent Equations of Single Differences

In this section, the equivalent equations are formed to eliminate the satellite clock errors from the original zero-difference equations first, then the equivalency of the single differences (in two cases) related to the original zero-difference equations is proved.

Single differences cancel all the satellite clock errors out of the observation equations. This can also be achieved by forming equivalent equations where satellite clock errors are eliminated. Considering (1), the original observation equation, and X_1, the vector of satellite clock errors, the equivalent equations of single differences can be formed as outlined in Sect. 7.2.2 (cf. Xu 2007).

Suppose n common satellites ($k1$,$k2$,. . .,kn) are observed at stations $i1$ and $i2$. The original observation equation can then be written as

$$\begin{pmatrix} V_{i1} \\ V_{i2} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} - \begin{pmatrix} E & B_{i1} \\ E & B_{i2} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad P = \frac{1}{\sigma^2}\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}, \tag{12}$$

where X_1 is the vector of satellite clock errors and X_2 is the vector of other unknowns. For simplicity, clock errors are scaled by the speed of light c and directly used as unknowns; then the X_1-related coefficient matrix is an identity matrix, E.

Comparing (12) with (1), one has (cf. Sect. 7.2.2)

$$A = \begin{pmatrix} E \\ E \end{pmatrix}, \quad B = \begin{pmatrix} B_{i1} \\ B_{i2} \end{pmatrix}, \quad L = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} V_{i1} \\ V_{i2} \end{pmatrix},$$

and

$$M_{11} = \begin{pmatrix} E & E \end{pmatrix}\frac{1}{\sigma^2}\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}\begin{pmatrix} E \\ E \end{pmatrix} = \frac{2}{\sigma^2}E\,,$$

$$J = \begin{pmatrix} E \\ E \end{pmatrix} \frac{\sigma^2}{2} E \begin{pmatrix} E & E \end{pmatrix} P = \frac{1}{2} \begin{pmatrix} E & E \\ E & E \end{pmatrix},$$

$$E_{2n\times 2n} - J = \frac{1}{2} \begin{pmatrix} E & -E \\ -E & E \end{pmatrix} \quad \text{and}$$

$$(E_{2n\times 2n} - J)B = \frac{1}{2} \begin{pmatrix} B_{i1} - B_{i2} \\ B_{i2} - B_{i1} \end{pmatrix}.$$

So the equivalently eliminated equation system of (12) is

$$\begin{pmatrix} U_{i1} \\ U_{i2} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} B_{i1} - B_{i2} \\ B_{i2} - B_{i1} \end{pmatrix} \cdot X_2, \quad P = \frac{1}{\sigma^2} \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}, \tag{13}$$

where the satellite clock error vector X_1 is eliminated, and the observable vector and weight matrix are unchanged.

Denoting $B_s = B_{i2} - B_{i1}$, the least squares normal equation of (13) can then be formed as (cf. Chap. 7 of Xu 2007) (supposing (13) is solvable)

$$\frac{1}{2} \begin{pmatrix} -B_s^T & B_s^T \end{pmatrix} \cdot P \cdot \begin{pmatrix} -B_s \\ B_s \end{pmatrix} \cdot X_2 = \begin{pmatrix} -B_s^T & B_s^T \end{pmatrix} \cdot P \cdot \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix}$$

or

$$B_s^T B_s \cdot X_2 = B_s^T (L_{i2} - L_{i1}). \tag{14}$$

Alternatively, a single-difference equation can be obtained by multiplying (12) with a transformation matrix C_s:

$$C_s = \begin{pmatrix} -E & E \end{pmatrix},$$

giving

$$C_s \cdot \begin{pmatrix} V_{i1} \\ V_{i2} \end{pmatrix} = C_s \cdot \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} - C_s \cdot \begin{pmatrix} E & B_{i1} \\ E & B_{i2} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

or

$$V_{i2} - V_{i1} = (L_{i2} - L_{i1}) - (B_{i2} - B_{i1}) X_2 \tag{15}$$

and

$$\text{cov}(\text{SD}(O)) = C_s \sigma^2 \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} C_s^T = 2\sigma^2 E \quad \text{and} \quad P_s = \frac{1}{2\sigma^2} E, \tag{16}$$

where P_s is the weight matrix of single differences, and cov(SD(O)) is the covariance of the single differences (SD) observational vector (O). Supposing (15) is solvable, the least squares normal equation system of (15) is then

$$(B_{i2} - B_{i1})^T (B_{i2} - B_{i1}) X_2 = (B_{i2} - B_{i1})^T (L_{i2} - L_{i1}). \tag{17}$$

It is obvious that (17) and (14) are identical. Therefore in the case of two stations, the single-difference equation (15) is equivalent to the equivalently eliminated equation (13) and consequently equivalent to the original zero-difference equation.

Suppose n common satellites ($k1$, $k2$, …, kn) are observed at stations $i1$, $i2$ and $i3$. The original observation equation can then be written as

$$\begin{pmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \\ L_{i3} \end{pmatrix} - \begin{pmatrix} E & B_{i1} \\ E & B_{i2} \\ E & B_{i3} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad P = \frac{1}{\sigma^2} \begin{pmatrix} E & 0 & 0 \\ 0 & E & 0 \\ 0 & 0 & E \end{pmatrix}. \tag{18}$$

Comparing (18) with (1), one has (cf. Sect. 7.2.2)

$$A = \begin{pmatrix} E \\ E \\ E \end{pmatrix}, \quad B = \begin{pmatrix} B_{i1} \\ B_{i2} \\ B_{i3} \end{pmatrix}, \quad L = \begin{pmatrix} L_{i1} \\ L_{i2} \\ L_{i3} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{pmatrix},$$

and

$$M_{11} = A^{\mathrm{T}} P A = \frac{3}{\sigma^2} E,$$

$$J = A \frac{\sigma^2}{3} E A^{\mathrm{T}} P = \frac{1}{3} \begin{pmatrix} E & E & E \\ E & E & E \\ E & E & E \end{pmatrix},$$

$$E_{3n\times 3n} - J = \frac{1}{3} \begin{pmatrix} 2E & -E & -E \\ -E & 2E & -E \\ -E & -E & 2E \end{pmatrix} \quad \text{and}$$

$$(E_{3n\times 3n} - J)B = \frac{1}{3} \begin{pmatrix} 2B_{i1} - B_{i2} - B_{i3} \\ -B_{i1} + 2B_{i2} - B_{i3} \\ -B_{i1} - B_{i2} + 2B_{i3} \end{pmatrix}.$$

So the equivalently eliminated equation system of (18) is

$$\begin{pmatrix} U_{i1} \\ U_{i2} \\ U_{i3} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \\ L_{i3} \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 2B_{i1} - B_{i2} - B_{i3} \\ -B_{i1} + 2B_{i2} - B_{i3} \\ -B_{i1} - B_{i2} + 2B_{i3} \end{pmatrix} \cdot X_2, \quad P = \frac{1}{\sigma^2} \begin{pmatrix} E & 0 & 0 \\ 0 & E & 0 \\ 0 & 0 & E \end{pmatrix}, \tag{19}$$

and the related least squares normal equation can be formed as

$$\begin{aligned} &\frac{1}{3} \begin{pmatrix} 2B_{i1} - B_{i2} - B_{i3} \\ -B_{i1} + 2B_{i2} - B_{i3} \\ -B_{i1} - B_{i2} + 2B_{i3} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2B_{i1} - B_{i2} - B_{i3} \\ -B_{i1} + 2B_{i2} - B_{i3} \\ -B_{i1} - B_{i2} + 2B_{i3} \end{pmatrix} X_2 \\ &= \begin{pmatrix} 2B_{i1} - B_{i2} - B_{i3} \\ -B_{i1} + 2B_{i2} - B_{i3} \\ -B_{i1} - B_{i2} + 2B_{i3} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} L_{i1} \\ L_{i2} \\ L_{i3} \end{pmatrix}. \end{aligned} \tag{20}$$

Alternatively, for the equation system (18), single differences can be formed using transformation (cf. Sect. 6.6.1 of Xu 2007):

$$C_{\mathrm{s}} = \begin{pmatrix} -E & E & 0 \\ 0 & -E & E \end{pmatrix}$$

and

$$P_{\mathrm{s}} = [\mathrm{cov}(\mathrm{SD})]^{-1} = \frac{1}{3\sigma^2} \begin{pmatrix} 2E & E \\ E & 2E \end{pmatrix}.$$

The correlation problem appears in the case of single differences of multiple baselines. The related observation equations and the least squares normal equation can be written as

$$\begin{pmatrix} V_{i2} - V_{i1} \\ V_{i3} - V_{i2} \end{pmatrix} = \begin{pmatrix} L_{i2} - L_{i1} \\ L_{i3} - L_{i2} \end{pmatrix} - \begin{pmatrix} B_{i2} - B_{i1} \\ B_{i3} - B_{i2} \end{pmatrix} X_2, \quad P_{\mathrm{s}} \quad \text{and} \tag{21}$$

$$\begin{aligned} &\begin{pmatrix} B_{i2} - B_{i1} \\ B_{i3} - B_{i2} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2E & E \\ E & 2E \end{pmatrix} \begin{pmatrix} B_{i2} - B_{i1} \\ B_{i3} - B_{i2} \end{pmatrix} X_2 \\ &= \begin{pmatrix} B_{i2} - B_{i1} \\ B_{i3} - B_{i2} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2E & E \\ E & 2E \end{pmatrix} \begin{pmatrix} L_{i2} - L_{i1} \\ L_{i3} - L_{i2} \end{pmatrix}. \end{aligned} \tag{22}$$

Equations (20) and (22) are identical. This may be proved by expanding both equations and comparing the results. Again, this shows that the equivalently eliminated equations are equivalent to the single-difference equations, however, without the need to deal with the correlation problem.

7.2.4 Equivalent Equations of Double Differences

Double differences cancel all the clock errors out of the observation equations. This can also be achieved by forming equivalent equations where all clock errors are eliminated. Considering (1), the original observation equation, and X_1, the vector of all clock errors, the equivalent equation of double differences can be formed as outlined in Sect. 7.2.2.

In the case of two stations, supposing n common satellites ($k1$, $k2$, …, kn) are observed at stations $i1$ and $i2$, the equivalent single-difference observation equation is then (13). Denoting $B_{\mathrm{s}1} = B_{i2} - B_{i1}$, the station clock error parameter as $\delta t_{i1} - \delta t_{i2}$ (cf. (6.89), (6.90), (6.91) and (6.92) of Xu 2007) and assigning the coefficients of the first column to the station clock errors, i.e. $B_{\mathrm{s}1} = (I_{n\times 1}\ B_{\mathrm{s}})$, (13) turns out to be

$$\begin{pmatrix} U_{i1} \\ U_{i2} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -I_{n\times 1} & -B_s \\ I_{n\times 1} & B_s \end{pmatrix} \begin{pmatrix} X_c \\ X_3 \end{pmatrix} \quad \text{and} \quad P = \frac{1}{\sigma^2} \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}, \tag{23}$$

where X_c is the station clock error vector, X_3 is the other unknown vector, B_s is the X_3-related coefficient matrix, $I_{n\times 1}$ is a 1 matrix (where all elements are 1), and clock errors are scaled by the speed of light.

Comparing (23) with (1), one has (cf. Sect. 7.2.2)

$$A = \frac{1}{2}\begin{pmatrix} -I_{n\times 1} \\ I_{n\times 1} \end{pmatrix}, \quad B = \frac{1}{2}\begin{pmatrix} -B_s \\ B_s \end{pmatrix}, \quad L = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} U_{i1} \\ U_{i2} \end{pmatrix},$$

and

$$M_{11} = \frac{1}{4}\left(-I_{n\times 1}^{\mathrm{T}} \quad I_{n\times 1}^{\mathrm{T}}\right)\frac{1}{\sigma^2}\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}\begin{pmatrix} -I_{n\times 1} \\ I_{n\times 1} \end{pmatrix} = \frac{n}{2\sigma^2},$$

$$J = \begin{pmatrix} -I_{n\times 1} \\ I_{n\times 1} \end{pmatrix}\frac{\sigma^2}{2n}\left(-I_{n\times 1}^{\mathrm{T}} \quad I_{n\times 1}^{\mathrm{T}}\right)\cdot P = \frac{1}{2n}\begin{pmatrix} I_{n\times n} & -I_{n\times n} \\ -I_{n\times n} & I_{n\times n} \end{pmatrix} \quad \text{and}$$

$$(E_{2n\times 2n} - J)\frac{1}{2}\begin{pmatrix} -B_s \\ B_s \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -E_{n\times n} + \dfrac{1}{n}I_{n\times n} \\ E_{n\times n} - \dfrac{1}{n}I_{n\times n} \end{pmatrix} B_s.$$

So the equivalently eliminated equation system of (23) is

$$\begin{pmatrix} U_{i1} \\ U_{i2} \end{pmatrix} = \begin{pmatrix} L_{i1} \\ L_{i2} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -E_{n\times n} + \dfrac{1}{n}I_{n\times n} \\ E_{n\times n} - \dfrac{1}{n}I_{n\times n} \end{pmatrix} B_s X_3 \quad \text{and} \quad P = \frac{1}{\sigma^2}\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}, \tag{24}$$

where the receiver clock error vector X_c is eliminated, and observable vector and weight matrix are unchanged. The normal equation has a simple form of

$$B_s^{\mathrm{T}}\left(E_{n\times n} - \frac{1}{n}I_{n\times n}\right)B_s X_3 = B_s^{\mathrm{T}}\left(E_{n\times n} - \frac{1}{n}I_{n\times n}\right)(L_{i2} - L_{i1}). \tag{25}$$

Alternatively, the traditional single-difference observation equations (15) and (16) can be rewritten as

$$V_{i2} - V_{i1} = (L_{i2} - L_{i1}) - \begin{pmatrix} I_{n\times 1} & B_s \end{pmatrix}\begin{pmatrix} X_c \\ X_3 \end{pmatrix}$$

or

$$\begin{pmatrix} V_{i2}^1 - V_{i1}^1 \\ V_{i2}^k - V_{i1}^k \end{pmatrix} = \begin{pmatrix} L_{i2}^1 - L_{i1}^1 \\ L_{i2}^k - L_{i1}^k \end{pmatrix} - \begin{pmatrix} 1 & B_s^1 \\ I_{m\times 1} & B_s^k \end{pmatrix}\begin{pmatrix} X_c \\ X_3 \end{pmatrix}, \tag{26}$$

and

$$\mathrm{cov}(\mathrm{SD}(O)) = C_s\sigma^2\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}C_s^{\mathrm{T}} = 2\sigma^2 E \quad \text{and} \quad P_s = \frac{1}{2\sigma^2}E,$$

where $m = n-1$, and the superscripts 1 and k denote the first row and remaining rows of the matrices (or columns in case of vectors). The double-difference transformation matrix and covariance are (cf. (6.116)–(6.118) of Xu 2007)

$$C_{\mathrm{d}} = \left(-I_{m\times 1} \; E_{m\times m} \right),$$

$$\mathrm{cov}(\mathrm{DD}(O)) = C_{\mathrm{d}}\,\mathrm{cov}(\mathrm{SD}(O))C_{\mathrm{d}}^{\mathrm{T}} = 2\sigma^2 C_{\mathrm{d}}C_{\mathrm{d}}^{\mathrm{T}} = 2\sigma^2\,(I_{m\times m} + E_{m\times m}) \text{ and}$$

$$P_{\mathrm{d}} = [\mathrm{cov}(\mathrm{DD}(O))]^{-1} = \frac{1}{2\sigma^2 n}\,(nE_{m\times m} - I_{m\times m})\,.$$

The double-difference observation equation and related normal equation are

$$C_{\mathrm{d}} \begin{pmatrix} V_{i2}^1 - V_{i1}^1 \\ V_{i2}^k - V_{i1}^k \end{pmatrix} = C_{\mathrm{d}} \begin{pmatrix} L_{i2}^1 - L_{i1}^1 \\ L_{i2}^k - L_{i1}^k \end{pmatrix} - C_{\mathrm{d}} \begin{pmatrix} 1 & B_{\mathrm{s}}^1 \\ I_{m\times 1} & B_{\mathrm{s}}^k \end{pmatrix} \begin{pmatrix} X_{\mathrm{c}} \\ X_3 \end{pmatrix}$$

or

$$C_{\mathrm{d}} \begin{pmatrix} V_{i2}^1 - V_{i1}^1 \\ V_{i2}^k - V_{i1}^k \end{pmatrix} = C_{\mathrm{d}} \begin{pmatrix} L_{i2}^1 - L_{i1}^1 \\ L_{i2}^k - L_{i1}^k \end{pmatrix} - C_{\mathrm{d}} \begin{pmatrix} B_{\mathrm{s}}^1 \\ B_{\mathrm{s}}^k \end{pmatrix} X_3,$$

i.e.

$$C_{\mathrm{d}}(V_{i2} - V_{i1}) = C_{\mathrm{d}}(L_{i2} - L_{i1}) - C_{\mathrm{d}}B_{\mathrm{s}}X_3 \tag{27}$$

and

$$B_{\mathrm{s}}^{\mathrm{T}}C_{\mathrm{d}}^{\mathrm{T}}P_{\mathrm{d}}C_{\mathrm{d}}B_{\mathrm{s}}X_3 = B_{\mathrm{s}}^{\mathrm{T}}C_{\mathrm{d}}^{\mathrm{T}}P_{\mathrm{d}}C_{\mathrm{d}}(L_{i2} - L_{i1}), \tag{28}$$

where

$$C_{\mathrm{d}}^{\mathrm{T}}P_{\mathrm{d}}C_{\mathrm{d}} = \frac{1}{2\sigma^2 n} \left(-I_{m\times 1} \quad E_{m\times m} \right)^{\mathrm{T}} (nE_{m\times m} - I_{m\times m}) \left(-I_{m\times 1} \quad E_{m\times m} \right), \tag{29}$$

$$\left(-I_{m\times 1} \quad E_{m\times m} \right)^{\mathrm{T}} (nE_{m\times m} - I_{m\times m}) = \left(-I_{m\times 1} \quad nE_{m\times m} - I_{m\times m} \right)^{\mathrm{T}} \tag{30}$$

and

$$\left(-I_{m\times 1} \quad nE_{m\times m} - I_{m\times m} \right)^{\mathrm{T}} \left(-I_{m\times 1} \quad E_{m\times m} \right) = nE_{n\times n} - I_{n\times n}. \tag{31}$$

The above three equations can be proved readily. Substituting (29), (30) and (31) into (28), (28) turns out to be the same as (25). So the equivalency between the double-difference equation and the directly formed equivalent equation (23) is proved.

7.2.5 Equivalent Equations of Triple Differences

Triple differences cancel all the clock errors and real-valued ambiguities out of the observation equations. This can also be achieved by forming equivalent equations where all clock errors and real-valued ambiguities are eliminated. Considering (1),

the original observation equation, and X_1, the parameter vector of all clock errors and real-valued ambiguities, the equivalent equations of triple differences can then be formed as outlined in Sect. 7.2.2.

It is well known that traditional triple differences are correlated between adjacent epochs and between baselines. In the case of sequential (epoch by epoch) data processing of triple differences, the correlation problem is difficult to deal with. However, using the equivalently eliminated equations, the weight matrix remains diagonal. The GPS observables remain the original ones.

7.2.6 *Method of Dealing with the Reference Parameters*

In differential GPS data processing, the reference-related parameters are usually considered known and are fixed (or not adjusted). This may be realised by the a priori datum method (for details cf. Sect. 7.8.2 of Xu 2007). Here we just outline the basic principle.

The equivalent observation equation system (9) can be rewritten as

$$U = L - \begin{pmatrix} D_1 & D_2 \end{pmatrix} \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix} \quad \text{and} \quad P\,, \tag{32}$$

where

$$D = \begin{pmatrix} D_1 & D_2 \end{pmatrix} \text{ and } X_2 = \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix}.$$

Suppose there are a priori constraints of (cf., e.g. Xu 2007)

$$W = \overline{X}_{22} - X_{22} \quad \text{and} \quad P_2\,, \tag{33}$$

where $\overline{X}_{22}$ is the "directly observed" parameter sub-vector, P_2 is the weight matrix with respect to the parameter sub-vector X_{22} and W is a residual vector, which has the same property as U. Usually, $\overline{X}_{22}$ is "observed" independently, so P_2 is a diagonal matrix. If X_{22} is a sub-vector of station coordinates, then the constraint of (33) is called a datum constraint. (This is also the reason why the name a priori datum is used.) We consider here X_{22}, a vector of reference-related parameters (such as clock errors and ambiguities of the reference satellite and reference station). Generally, the a priori weight matrix P_2 is given by covariance matrix Q_{W} and

$$P_2 = Q_{\mathrm{W}}^{-1}. \tag{34}$$

In practice, the sub-vector $\overline{X}_{22}$ is usually a zero vector; this can be achieved through careful initialisation by forming observation equation (1).

The least squares normal equation of the a priori datum problem of (32) and (33) can be formed. Compared with the normal equation of (32), the only difference between the two normal equations is that the a priori weight matrix P_2 has been added to the normal matrix. This indicates that the a priori datum problem can be dealt with simply by adding P_2 to the normal equation of observation equation (32).

If some diagonal components of the weight matrix P_2 are set to zero, then the related parameters (in X_{22}) are free parameters (or free datum) of the adjustment problem (without a priori constraints). Otherwise, parameters with a priori constraints are called a priori datum. Large weight indicates strong constraint and small weight indicates soft constraint. The strongest constraint is to keep the datum fixed. The reference-related datum (coordinates and clock errors as well as ambiguities) can be fixed by applying the strongest constraints to the related parameters, i.e. by adding the strongest constraints to the datum-related diagonal elements of the normal matrix.

7.2.7 Summary of the Unified Equivalent Algorithm

For any linearised zero-difference GPS observation equation system (1)

$$V = L - \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad P\,, \tag{35}$$

the X_1-eliminated equivalent GPS observation equation system is then (9)

$$U_2 = L - (E - J)BX_2 \quad \text{and} \quad P, \tag{36}$$

where

$$J = AM_{11}^{-1}A^{\mathrm{T}}P\,, \qquad M_{11} = A^{\mathrm{T}}PA, \tag{37}$$

where E is an identity matrix, L is original observational vector, P is original weight matrix and U_2 is residual vector, which has the same property as V.

Similarly, the X_2-eliminated equivalent equation system is (11)

$$U_1 = L - (E - K)AX_1 \quad \text{and} \quad P\,, \tag{38}$$

where

$$K = BM_{22}^{-1}B^{\mathrm{T}}P\,, \qquad M_{22} = B^{\mathrm{T}}PB \tag{39}$$

and U_1 is the residual vector (which has the same property as V).

The fixing of the values of sub-vector X_{22} (of X_2) can be realised by adding the strongest constraints to the X_{22}-related diagonal elements of the normal matrix formed by (36). Alternatively, we may apply the strongest constraints directly to the normal equation formed by (35) first. In this way, the reference-related parameters (clock errors, ambiguities, coordinates, etc.) are fixed. And then we may form the equivalently eliminated observation equation (36). In this way, the relative and differential GPS data processing can be realised by using (36) after selecting X_1 to be eliminated.

The GPS data processing algorithm using (36) is then a selectively eliminated equivalent method. Selecting X_1 in (35) as a zero vector, the algorithm is identical to the zero-difference method. Selecting X_1 in (35) as the satellite clock error vector,

the vector of all clock errors, the clock error and ambiguity vector and any user-defined vector, the algorithm is equivalent to the single-difference method, double-difference method, triple-difference method and user-defined eliminating method, respectively. The eliminated unknown X_1 can be solved separately if desired.

The advantages of this method are (compared with un-differential and differential methods) the following:

- The un-differential and differential GPS data processing can be dealt with in an equivalent and unified way. The data processing scenarios can be selected by a switch and used in a combinative way.
- The eliminated parameters can be also solved separately with the same algorithm.
- The weight matrix remains the original diagonal one.
- The original observations are used; no differencing is required.

Meanwhile, it is obvious that the described algorithm has all the advantages of all un-differential and differential GPS data processing.

7.3 Equivalence of the Uncombined and Combining Algorithms

Uncombined and combining algorithms are standard GPS data processing methods, which can often be found in the literature (cf., e.g. Leick 2004; Hofmann-Wellenhof et al. 2001). Different combinations have different properties and are beneficial for dealing with the data and solving the problem in different cases (Hugentobler et al. 2001; Kouba and Heroux 2001; Zumberge et al. 1997). The equivalence between the undifferenced and differencing algorithms was proved and a unified equivalent data processing method was proposed by Xu (2002a, cf. Sect. 7.2). The question of whether the uncombined and combining algorithms are also equivalent is an interesting topic and will be addressed here in detail (cf. Xu et al. 2006a, 2007).

7.3.1 Uncombined GPS Data Processing Algorithms

7.3.1.1 Original GPS Observation Equations

The original GPS code pseudo-range and carrier-phase measurements represented in Xu (2007) ((6.44) and (6.45) in Sect. 7.6.5) can be simplified as

$$R_j = C_\rho + \delta_{\text{ion}}(j), \tag{40}$$

$$\lambda_j \Phi_j = C_\rho + \lambda_j N_j - \delta_{\text{ion}}(j), \quad j = 1, 2 \tag{41}$$

where

$$C_\rho = \rho - (\delta t_{\text{r}} - \delta t_{\text{k}})c + \delta_{\text{trop}} + \delta_{\text{tide}} + \delta_{\text{rel}} + \varepsilon_i, \quad i = c,\ p \tag{42}$$

$$\delta_{\text{ion}}(j) = \frac{A_1}{f_j^2} = \frac{A_1^z}{f_j^2} F = \frac{f_s^2 B_1}{f_j^2} = \frac{f_s^2 B_1^z}{f_j^2} F, \tag{43}$$

where R is the observed pseudo-range, Φ is the observed phase, c denotes the speed of light and δt_{r} and δt_{k} denote the clock errors of the receiver and satellite, respectively. The terms δ_{ion}, δ_{trop}, δ_{tide} and δ_{rel} denote the ionospheric, tropospheric, tidal and relativistic effects, respectively. Tidal effects include Earth tide effects and ocean-loading tide effects. The multipath effect is omitted here. The ρ is the geometric distance. All terms have units of length (meters). j is the index of the frequency f and wavelength λ. A_1 and $A_1{}^z$ are the ionospheric parameters in the path and zenith directions; B_1 and $B_1{}^z$ are scaled A_1 and $A_1{}^z$ with $f_s{}^2$ for numerical reason. Index c denotes code. C_ρ is called geometry and N_j is the ambiguity. For simplicity, the residuals of the codes (and phases) are denoted with the same symbol ε_{c} (and ε_{p}) and have the same standard deviations of σ_{c} (and σ_{p}). Equations (40) and (41) can be written in a matrix form with weight matrix P as (Blewitt 1998)

$$\begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 1 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 1 & -f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}, \; P = \begin{pmatrix} \sigma_{\text{c}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\text{c}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\text{p}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\text{p}}^2 \end{pmatrix}^{-1}. \tag{44}$$

7.3.1.2 Solutions of Uncombined Observation Equations

Equation (44) includes the observations of one satellite viewed by one receiver at one epoch. Alternatively, (44) can be considered as a transformation between the observations and unknowns and the transformation is a linear and invertible one. Denoting

$$a = \frac{f_1^2}{f_1^2 - f_2^2}, \quad b = \frac{-f_2^2}{f_1^2 - f_2^2}, \quad g = \frac{1}{f_1^2} - \frac{1}{f_2^2}, \quad q = g f_s^2, \tag{45}$$

one has relations of

$$1 - a = b, \quad \frac{1}{f_1^2\, g} = b, \quad \frac{1}{f_2^2\, g} = -a \tag{46}$$

and

$$\begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 1 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 1 & -f_s^2/f_2^2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q & -1/q & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} = T, \tag{47}$$

where a and b are the coefficients of the ionosphere-free combinations of the observables of L1 and L2. The solution of (44) has a form (multiplying (44) by the transformation matrix T) of

$$\begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q & -1/q & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1 \Phi_1 \\ \lambda_2 \Phi_2 \end{pmatrix}. \tag{48}$$

The related covariance matrix of the above solution vector is then

$$Q = \mathrm{cov} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix} = T \begin{pmatrix} \sigma_c^2 & 0 & 0 & 0 \\ 0 & \sigma_c^2 & 0 & 0 \\ 0 & 0 & \sigma_p^2 & 0 \\ 0 & 0 & 0 & \sigma_p^2 \end{pmatrix} T^{\mathrm{T}}$$

$$= \begin{pmatrix} (1-2a)^2 + 4b^2 + \dfrac{\sigma_p^2}{\sigma_c^2} & 4a^2 - 4ab - 2a + 2b & \dfrac{1-2a+2b}{q} & a - 2a^2 - 2b^2 \\ 4a^2 - 4ab - 2a + 2b & 8a^2 - 4a + 1 + \dfrac{\sigma_p^2}{\sigma_c^2} & \dfrac{1-4a}{q} & -2a^2 + 2ab - b \\ \dfrac{1-2a+2b}{q} & \dfrac{1-4a}{q} & \dfrac{2}{q^2} & \dfrac{a-b}{q} \\ a - 2a^2 - 2b^2 & -2a^2 + 2ab - b & \dfrac{a-b}{q} & a^2 + b^2 \end{pmatrix} \sigma_c^2. \tag{49}$$

Equation (49) can be simplified by using the relation of $1–a=b$ and neglecting the terms of $(\sigma_p/\sigma_c)^2$ (because (σ_p/σ_c) is less than 0.01) as well as letting $f_s=f_1$ (so that $q=1/b$). Taking the relationships of ratios of the frequencies into account ($f_1=154f_0$ and $f_2=120f_0$, f_0 is the fundamental frequency), one has approximately

$$\mathrm{cov} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix} = \begin{pmatrix} 26.2971 & 33.4800 & 11.1028 & -15.1943 \\ 33.4800 & 42.6629 & 14.1943 & -19.2857 \\ 11.1028 & 14.1943 & 4.7786 & -6.3243 \\ -15.1943 & -19.2857 & -6.3243 & 8.8700 \end{pmatrix} \sigma_c^2. \tag{50}$$

The precisions of the solutions will be further discussed in Sect. 7.3.3. The parameterisation of the GPS observation models is an important issue and can be found in Sect. 7.4 or Blewitt (1998) and Xu (2004) if interested.

7.3.2 Combining Algorithms of GPS Data Processing

The ionosphere-free combinations, geometry-free combinations, ionosphere-free and geometry-free combinations, diagonal combinations and wide-lane and narrow-lane combinations can be formed by multiplying the following transformation matrices to (44). The transformations and the transformed (i.e. combined) equations are listed below.

Ionosphere-free transformation and combinations are

$$T_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ a & b & 0 & 0 \\ 0 & 0 & a & b \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}, \quad \text{and}$$

$$T_1 \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1 \Phi_1 \\ \lambda_2 \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & q & 0 \\ 0 & 0 & 0 & 1 \\ a & b & 0 & 1 \\ 1/2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}. \tag{51}$$

The ionosphere parameter in (51) is free in the last three equations, which are traditionally called ionosphere-free combinations.

Geometry-free transformation and combinations are

$$T_2 = \begin{pmatrix} a & b & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 1 & 0 \end{pmatrix}, \quad \text{and}$$

$$T_2 \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1 \Phi_1 \\ \lambda_2 \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & q & 0 \\ 1 & -1 & -q & 0 \\ 1 & 0 & -2f_s^2/f_1^2 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}. \tag{52}$$

The geometric component in (52) is free in the last three equations, which are traditionally called geometry-free combinations.

Ionosphere-free and geometry-free transformation and combinations are

$$T_3 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ a & b & 0 & 0 \\ -a & -b & a & b \\ 1/2 - a & -b & 1/2 & 0 \end{pmatrix}, \quad \text{and} \quad T_3 \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1 \Phi_1 \\ \lambda_2 \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & q & 0 \\ 0 & 0 & 0 & 1 \\ a & b & 0 & 0 \\ 1/2 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}. \tag{53}$$

The ionosphere and geometry are both free in the last two equations, which are called ionosphere-geometry-free combinations.

Diagonal transformation and combinations are

$$T_4 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ a & b & 0 & 0 \\ -2ab & b(2a-1) & 0 & b \\ 1/2 - a & -b & 1/2 & 0 \end{pmatrix}, \quad \text{and} \quad T_4 \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1 \Phi_1 \\ \lambda_2 \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & q & 0 \\ 0 & 0 & 0 & 1 \\ 0 & b & 0 & 0 \\ 1/2 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}. \tag{54}$$

In the above equation, the ionosphere and geometry as well as the ambiguities are diagonal to each other. Such combinations are called diagonal ones.

Wide-lane and narrow-lane transformation and combinations are

$$T_5 = \begin{pmatrix} 0 & 0 & -1/b\lambda_2 & 2/\lambda_1 \\ 0 & 0 & 1/b\lambda_2 & 2/\lambda_1 \\ -1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} T_4, \quad \text{and} \quad T_5 \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix} = \begin{pmatrix} N_1 - N_2 \\ N_1 + N_2 \\ B_1 \\ C_\rho \end{pmatrix}.$$

7.3.2.1 General Combinations

For arbitrary combinations, as soon as the transformation matrix is an invertible one, the transformed equations are equivalent to the original ones based on algebra theory. Both the solution vector and the variance–covariance matrix are identical. That is, no matter what kinds of combinations are used, neither different solutions nor different precisions of the solutions will be obtained. Therefore, all the above combinations have the same results of (48). The different combinations lead to an easier dealing of the related special problems.

7.3.3 Secondary GPS Data Processing Algorithms

7.3.3.1 In the Case of More Satellites in View

Up to now, the discussions have been limited for the observations of one satellite viewed by one receiver at one epoch. The original observation equation is given in (44). The solution vector and its covariance matrix are given in (48) and (49), respectively. The elements of the covariance matrix depend on the coefficients of (44), and the coefficients of the observation equation depend on the way of parameterisation. For example, if instead of B_1, $B_1{}^z$ is used, then (44) turns out to be

$$\begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1\Phi_1(k) \\ \lambda_2\Phi_2(k) \end{pmatrix} = \begin{pmatrix} 0 & 0 & F_k f_s^2/f_1^2 & 1 \\ 0 & 0 & F_k f_s^2/f_2^2 & 1 \\ 1 & 0 & -F_k f_s^2/f_1^2 & 1 \\ 0 & 1 & -F_k f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1^z \\ C_\rho(k) \end{pmatrix}, \tag{55}$$

where k is the index of the satellite. Ionospheric mapping function F_k is dependent on the zenith distance of the satellite k. The solution vector of (55) is then similar to that of (48):

$$\begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1^z \\ C_\rho(k) \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q_k & -1/q_k & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} \begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1\Phi_1(k) \\ \lambda_2\Phi_2(k) \end{pmatrix}, \; Q(k), \tag{56}$$

where $q_k{=}qF_k$ and $Q(k)$ is the covariance matrix, which can be similarly derived and given by adding the index k to q in Q of (49). The terms on the right-hand side can be considered secondary "observations" of the unknowns on the left-hand side. If K satellites are viewed, one has the observation equations of one receiver:

$$\begin{pmatrix} \lambda_1 N_1(1) \\ \lambda_2 N_2(1) \\ B_1^z \\ C_\rho(1) \\ \vdots \\ \lambda_1 N_1(K) \\ \lambda_2 N_2(K) \\ B_1^z \\ C_\rho(K) \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1\,0 & \cdots & 0 & 0 & 0\,0 \\ -2a & 2a-1 & 0\,1 & \cdots & 0 & 0 & 0\,0 \\ 1/q_1 & -1/q_1 & 0\,0 & \cdots & 0 & 0 & 0\,0 \\ a & b & 0\,0 & \cdots & 0 & 0 & 0\,0 \\ \vdots & \vdots & \vdots\,\vdots & \vdots\,\vdots & \vdots & \vdots & \vdots\,\vdots \\ 0 & 0 & 0\,0 & \cdots & 1-2a & -2b & 1\,0 \\ 0 & 0 & 0\,0 & \cdots & -2a & 2a-1 & 0\,1 \\ 0 & 0 & 0\,0 & \cdots & 1/q_K & -1/q_K & 0\,0 \\ 0 & 0 & 0\,0 & \cdots & a & b & 0\,0 \end{pmatrix} \begin{pmatrix} R_1(1) \\ R_2(1) \\ \lambda_1 \Phi_1(1) \\ \lambda_2 \Phi_2(1) \\ \vdots \\ R_1(K) \\ R_2(K) \\ \lambda_1 \Phi_1(K) \\ \lambda_2 \Phi_2(K) \end{pmatrix}, \tag{57}$$

and variance matrix

$$Q_K = \begin{pmatrix} Q(1) & \cdots & 0 \\ \vdots & \ldots & \vdots \\ 0 & \cdots & Q(K) \end{pmatrix}. \tag{58}$$

Multiplying a transformation matrix

$$T(K) = \begin{pmatrix} 1\,0 & 0 & 0 & \cdots & 0\,0 & 0 & 0 \\ 0\,1 & 0 & 0 & \cdots & 0\,0 & 0 & 0 \\ 0\,0 & 1/K & 0 & \cdots & 0\,0 & 1/K & 0 \\ 0\,0 & 0 & 1 & \cdots & 0\,0 & 0 & 0 \\ \vdots\,\vdots & \vdots & \vdots & \vdots\,\vdots & \vdots\,\vdots & \vdots & \vdots \\ 0\,0 & 0 & 0 & \cdots & 1\,0 & 0 & 0 \\ 0\,0 & 0 & 0 & \cdots & 0\,1 & 0 & 0 \\ 0\,0 & 0 & 0 & \cdots & 0\,0 & 0 & 1 \end{pmatrix} \tag{59}$$

to (57), one has the solutions of GPS observation equations of one station:

$$\begin{pmatrix} \lambda_1 N_1(1) \\ \lambda_2 N_2(1) \\ B_1^z \\ C_\rho(1) \\ \vdots \\ \lambda_1 N_1(K) \\ \lambda_2 N_2(K) \\ C_\rho(K) \end{pmatrix} = T(K) \begin{pmatrix} 1-2a & -2b & 1\,0 & \cdots & 0 & 0 & 0\,0 \\ -2a & 2a-1 & 0\,1 & \cdots & 0 & 0 & 0\,0 \\ 1/q_1 & -1/q_1 & 0\,0 & \cdots & 0 & 0 & 0\,0 \\ a & b & 0\,0 & \cdots & 0 & 0 & 0\,0 \\ \vdots & \vdots & \vdots\,\vdots & \vdots\,\vdots & \vdots & \vdots & \vdots\,\vdots \\ 0 & 0 & 0\,0 & \cdots & 1-2a & -2b & 1\,0 \\ 0 & 0 & 0\,0 & \cdots & -2a & 2a-1 & 0\,1 \\ 0 & 0 & 0\,0 & \cdots & 1/q_K & -1/q_K & 0\,0 \\ 0 & 0 & 0\,0 & \cdots & a & b & 0\,0 \end{pmatrix} \begin{pmatrix} R_1(1) \\ R_2(1) \\ \lambda_1 \Phi_1(1) \\ \lambda_2 \Phi_2(1) \\ \vdots \\ R_1(K) \\ R_2(K) \\ \lambda_1 \Phi_1(K) \\ \lambda_2 \Phi_2(K) \end{pmatrix}, \tag{60}$$

and the related

$$Q = T(K)Q_K(T(K))^{\mathrm{T}}, \tag{61}$$

where mapping function is used to combine the K ionosphere parameters into one. Similar discussions can be made for the cases of using more receivers. The original observation vector and the so-called secondary observation vector are

$$\begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1\Phi_1(k) \\ \lambda_2\Phi_2(k) \end{pmatrix}, \quad \begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1(k) \\ C_\rho(k) \end{pmatrix}. \tag{62}$$

Both vectors are equivalent as proved in Sect. 7.3.2 and they can be transformed uniquely from one to the other. Any further data processing can be considered processing based on the secondary "observations". The secondary "observations" have the equivalence property whether they are uncombined or combining ones. Therefore the equivalence property is valid for further data processing based on the secondary "observations".

7.3.3.2 GPS Data Processing Using Secondary "Observations"

A by-product of the above equivalence discussions is that the GPS data processing can be performed directly by using the so-called secondary observations. Besides the two ambiguity parameters (scaled with the wavelengths), the other two secondary observations are the electronic density in the observing path (scaled by square of f_1) and the geometry. The geometry includes the whole observation model except the ionosphere and ambiguity terms. For a time series of the secondary "observations", the electron density (or, for simplicity, "ionosphere") and the "geometry" are real-time observations, whereas the "ambiguities" are constants in case no cycle-slip occurs (Langley 1998a, b). Sequential adjustment or filtering methods can be used to deal with the observation time series. It is notable that the secondary "observations" are correlated with each other (see the covariance matrix (49)). However, the "ambiguities" are direct observations of the ambiguity parameters, and the "ionosphere" and "geometry" are modelled by (42) and (43), respectively. The "ambiguity" observables are ionosphere-geometry free. The "ionosphere" observable is geometry-free and ambiguity-free. The "geometry" observable is ionosphere-free. It is notable that some algorithms may be more effective; however, the results and the precisions of the solutions are equivalent no matter which algorithms are used. It should be emphasised that all the above discussions are based on the observation model (44). The problem concerning the parameterisation of the GPS observation model will not affect the conclusions of the discussions and will be further discussed in Sect. 7.4.

7.3.3.3 Precision Analysis

If the sequential time series of the original observations is considered time-independent as they traditionally have been, then the secondary "observations" and their precisions are also independent time series. From (50), the standard deviations of the L1 and L2 ambiguities are approximately $5.1281\sigma_c$ and $6.5317\sigma_c$. The standard deviations of ionosphere and geometry "observations" are about $2.1860\sigma_c$ and $2.9783\sigma_c$, respectively. That is, the precisions of the "observed" ambiguities are worse than those of the others at one epoch. If the standard deviation of the P code is about 1 dm (phase smoothed), then the precisions of the ambiguities determined by one epoch are worse than 0.5 m. However, an average filter of m epoch data will raise the precisions by a factor of sqrt(m) (square root of m). After 100 or 10,000 epochs, the ambiguities can be determined with precisions of about 5 cm or 5 mm. "Ionosphere" data are observed with better precision. However, due to the high dynamics of the electron movements, ionosphere effects may not be easily smoothed to raise the precision. The "geometry" model is the most complicated one compared with the others, and discussions can be found in numerous publications for static, kinematic and dynamic applications (cf., e.g. ION proceedings, Chap. 10 of Xu 2007).

7.3.4 Summary of the Combining Algorithms

The equivalence properties between uncombined and combining algorithms are proved theoretically by algebraic linear transformations. The solution vector and the related covariance matrix are identical, no matter which algorithms are used. Different combinations can lead to a more effective and easier dealing with the data. The so-called ionosphere-geometry-free combination and diagonal combination are derived, which have better properties than those of the traditional combinations. A data processing algorithm using the uniquely transformed secondary "observations" is outlined and used to prove the equivalence. Because of the unique property of the solutions of different combinations, any direct combinations of the solutions must be equivalent to each other. None of the combinations will lead to better solutions or better precisions of the solutions than those of the others. From this aspect, the traditional wide-lane ambiguity fixing technique may lead to a more effective search of ambiguity, but it will not lead to a better solution and precision of the ambiguity.

7.4 Parameterisation of the GPS Observation Model

The equivalences of combining and differencing algorithms are discussed in Sects. 7.3 and 7.2, respectively. The equivalence of the combining methods is an exact one, whereas the equivalence of the differencing algorithms is slightly

different. The parameters are implicitly expressed in the discussions; therefore, the parameterisation problems of the equivalent methods have not been discussed in detail. At that time, this topic was considered one of the remaining GPS theoretical problems (Xu 2003b, pp. 279–280; Wells et al. 1987, p. 34), and it will be discussed in the next section.

Three pieces of evidence of the parameterisation problem of the undifferenced GPS observation model are given first. Then the theoretical analysis and numerical derivation are made to show how to parameterise the bias effects of the undifferenced GPS observation model independently. A geometry-free illustration and a correlation analysis in the case of a phase–code combination are made. At the end, conclusions and comments are given.

7.4.1 Evidence of the Parameterisation Problem of the Undifferenced Observation Model

7.4.1.1 Evidence from Undifferenced and Differencing Algorithms

Suppose that the undifferenced GPS observation equation and the related LS normal equation are

$$V = L - \begin{pmatrix} A_1 & A_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad P \tag{63}$$

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \tag{64}$$

where all symbols have the same meanings as that of (1) and (5). Equation (64) can be diagonalised as (cf. Sect. 7.6.1 of Xu 2007)

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}. \tag{65}$$

The related equivalent observation equation of the diagonal normal equation (65) can be written as

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} L \\ L \end{pmatrix} - \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}, \tag{66}$$

where all symbols have the same meanings as that of (7.142) and (7.140) of Xu (2007). If X_1 is the vector containing all clock errors, then the second equation of (65) is the equivalent double differencing GPS normal equation. It is well known that in a double differencing algorithm, the ambiguity sub-vector contained in X_2 must be the double differencing ambiguities; otherwise, the problem will be generally singular. It is notable that X_2 is identical with that of the original undifferenced observation equation (63). Therefore, the ambiguity sub-vector contained in X_2 (in (63)) must be a set of double differencing ambiguities (or an equivalent set of

ambiguities). This is the first piece of evidence (or indication) of the singularity of the undifferenced GPS observation model in which the undifferenced ambiguities are used.

7.4.1.2 Evidence from Uncombined and Combining Algorithms

Suppose the original GPS observation equation of one viewed satellite is (cf. (44))

$$\begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 1 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 1 & -f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}, \; P; \tag{67}$$

then the uncombined or combining algorithms have the same solution of (cf. (48))

$$\begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q & -1/q & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix}, \tag{68}$$

where all symbols have the same meanings as that of (44) and (48). Then one notices that the ionosphere (B_1) and geometry (C_ρ) are functions of the codes (R_1 and R_2) and are independent from phases (Φ_1 and Φ_2) in (68). In other words, the phase observables do not have any contribution to the ionosphere and geometry. And this is not possible. Such an illogical conclusion is caused by the parameterisation of the ambiguities given in the observation model of (67). If one takes the first evidence discussed above into account and defines that, for each station, one of the satellites in view must be selected as reference and the related ambiguity has to be merged into the clock parameter, then the phases do have contributions to ionosphere and geometry. One notices again that the parameterisation is a very important topic and has to be discussed more specifically. An improper parameterisation of the observation model will lead to incorrect conclusions through the derivation from the model.

7.4.1.3 Evidence from Practice

Without using a priori information, a straightforward programming of the GPS data processing using an undifferenced algorithm leads to no results (i.e. the normal equation is singular, cf. Xu 2004). Therefore an exact parameterisation description is necessary and will be discussed in the next section.

7.4.2 A Method of Uncorrelated Bias Parameterisation

We restrict ourselves here to discussion of the parameterisation problem of the bias parameters (or constant effects, i.e. the clock errors and ambiguities) only.

Recalling the discussions of the equivalence of undifferenced and differencing algorithms in Sect. 7.2, the equivalence property is valid under three conditions: observation vector L used in (63) is identical; parameterisation of X_2 is identical and X_1 is able to be eliminated.

The first condition is necessary for the exactness of the equivalence because of the fact that, through forming differences, the unpaired data will be cancelled out in the differencing.

The second condition states that the parameterisation of the undifferenced and differencing model should be the same. This may be interpreted as the following: the rank of the undifferenced and differencing equations should be the same if the differencing is formed by a full rank linear transformation. If only the differencing equations are taken into account, then the rank of the undifferenced model should equal the rank of the differencing model plus the number of eliminated independent parameters.

It is well known that one of the clock error parameters is linearly correlated with the others. This may be seen in the proof of the equivalence property of the double differences, where the two receiver clock errors of the baseline may not be separated from each other and have to be transformed to one parameter and then eliminated (cf. Sect. 2 or Xu 2002a). This indicates that, if in the undifferenced model all clock errors are modelled, the problem will be singular (i.e. rank defect). Indeed, Wells et al. (1987) noticed that the equivalence is valid if measures are taken to avoid rank defect in the bias parameterisation. Which clock error has to be kept fixed is arbitrary. Because of the different qualities of the satellite and receiver clocks, a good choice is to fix a satellite clock error (the clock is called a reference clock). In practice, the clock error is an unknown; therefore there is no way to keep that fixed except to fix it to zero. In such a case, the meaning of the other bias parameters will be changed and may represent the relative errors between the other biases.

The third condition is important to ensure a full-ranked parameterisation of the parameter vector X_1 which is going to be eliminated.

The undifferenced equation (63) is solvable if the parameters X_1 and X_2 are not over-parameterised. In the case of single differences, X_1 includes satellite clock errors and is able to be eliminated. Therefore, to guarantee that the undifferenced model (63) is not singular, X_2 in (63) must not be over-parameterised. In the case of double differences, X_1 includes all clock errors except the reference one. Here we notice that the second observation equation of (63) is equivalent to the double differencing observation equation and the second equation of (64) is the related normal equation. In a traditional double differencing observation equation, the ambiguity parameters are represented by double differencing ambiguities. Recall that, for the equivalence property, the number (or rank) of ambiguity parameters in X_2 that are not linearly correlated must be equal to the number of the double differencing ambiguities. In the case of triple differences, X_1 includes all clock errors and ambiguities. The fact that X_1 should be able to be eliminated leads again to the conclusion that the ambiguities should be linearly independent.

The two equivalent linear equations should have the same rank. Therefore, if all clock errors except the reference one are modelled, the number of independent undifferenced ambiguity parameters should be equal to the number of double differencing ambiguities. According to the definition of the double differencing ambiguity, one has for one baseline

$$\begin{aligned}
N_{i1,i2}^{k1,k2} &= N_{i2}^{k2} - N_{i1}^{k2} - N_{i2}^{k1} + N_{i1}^{k1}, \\
N_{i1,i2}^{k1,k3} &= N_{i2}^{k3} - N_{i1}^{k3} - N_{i2}^{k1} + N_{i1}^{k1}, \\
N_{i1,i2}^{k1,k4} &= N_{i2}^{k4} - N_{i1}^{k4} - N_{i2}^{k1} + N_{i1}^{k1}, \\
&\vdots \\
N_{i1,i2}^{k1,kn} &= N_{i2}^{kn} - N_{i1}^{kn} - N_{i2}^{k1} + N_{i1}^{k1},
\end{aligned} \tag{69}$$

where $i1$ and $i2$ are station indices, kj is the jth satellite's identification, n is the common observed satellite number and is a function of the baseline and N is ambiguity. Then there are n–1 double differencing ambiguities and $2n$ undifferenced ambiguities. Taking the connection of the baselines into account, there are n–1 double differencing ambiguities and n new undifferenced ambiguities for any further baseline. If $i1$ is defined as the reference station of the whole network and $k1$ as the reference satellite of station $i2$, then undifferenced ambiguities of the reference station cannot be separated from the others (i.e. they are linearly correlated with the others). The undifferenced ambiguity of the reference satellite of station $i2$ cannot be separated from the others (i.e. it is linearly correlated with the others). That is, the ambiguities of the reference station cannot be determined, and the ambiguities of the reference satellites of non-reference stations cannot be determined. Either they should not be modelled or they should be kept fixed. A straightforward parameterisation of all undifferenced ambiguities will lead to rank defect and the problem will be singular and not able to be solved.

Therefore, using the equivalence properties of the equivalent equation of GPS data processing, we come to the conclusion that the ambiguities of the reference station and ambiguities of the reference satellite of every station are linearly correlated with the other ambiguities and clock error parameters. However, a general method of parameterisation should be independent of the selection of the references (station and satellite). Therefore, we use a two-baseline network to further our analysis. The original observation equation can be written as follows:

$$\begin{aligned}
L_{i1}^{k1} &= \cdots \delta_{i1} + \delta_{k1} + N_{i1}^{k1} + \cdots, \\
L_{i1}^{k2} &= \cdots \delta_{i1} + \delta_{k2} + N_{i1}^{k2} + \cdots, \\
L_{i1}^{k3} &= \cdots \delta_{i1} + \delta_{k3} + N_{i1}^{k3} + \cdots, \\
L_{i1}^{k4} &= \cdots \delta_{i1} + \delta_{k4} + N_{i1}^{k4} + \cdots, \\
L_{i1}^{k5} &= \cdots \delta_{i1} + \delta_{k5} + N_{i1}^{k5} + \cdots, \\
L_{i1}^{k6} &= \cdots \delta_{i1} + \delta_{k6} + N_{i1}^{k6} + \cdots,
\end{aligned} \tag{70}$$

$$
\begin{aligned}
L_{i2}^{k1} &= \cdots \delta_{i2} + \delta_{k1} + N_{i2}^{k1} + \cdots, \\
L_{i2}^{k2} &= \cdots \delta_{i2} + \delta_{k2} + N_{i2}^{k2} + \cdots, \\
L_{i2}^{k3} &= \cdots \delta_{i2} + \delta_{k3} + N_{i2}^{k3} + \cdots, \\
L_{i2}^{k4} &= \cdots \delta_{i2} + \delta_{k4} + N_{i2}^{k4} + \cdots, \\
L_{i2}^{k5} &= \cdots \delta_{i2} + \delta_{k5} + N_{i2}^{k5} + \cdots, \\
L_{i2}^{k7} &= \cdots \delta_{i2} + \delta_{k7} + N_{i2}^{k7} + \cdots,
\end{aligned} \tag{71}
$$

$$
\begin{aligned}
L_{i3}^{k2} &= \cdots \delta_{i3} + \delta_{k2} + N_{i3}^{k2} + \cdots, \\
L_{i3}^{k3} &= \cdots \delta_{i3} + \delta_{k3} + N_{i3}^{k3} + \cdots, \\
L_{i3}^{k4} &= \cdots \delta_{i3} + \delta_{k4} + N_{i3}^{k4} + \cdots, \\
L_{i3}^{k5} &= \cdots \delta_{i3} + \delta_{k5} + N_{i3}^{k5} + \cdots, \\
L_{i3}^{k6} &= \cdots \delta_{i3} + \delta_{k6} + N_{i3}^{k6} + \cdots, \\
L_{i3}^{k7} &= \cdots \delta_{i3} + \delta_{k7} + N_{i3}^{k7} + \cdots,
\end{aligned} \tag{72}
$$

where only the bias terms are listed and L and δ represent observable and clock error, respectively. Observation equations of stations $i1$, $i2$ and $i3$ are (70), (71) and (72). Define that the baselines 1, 2 are formed by stations $i1$ and $i2$, as well as by $i2$ and $i3$, respectively. Select $i1$ as the reference station and then keep the related ambiguities fixed (set to zero for simplification). For convenience of later discussion, select δ_{i1} as the reference clock (set to zero, too) and select $k1$, $k2$ as reference satellites of the stations $i2$, $i3$ (set the related ambiguities to zero), respectively. Then (70), (71) and (72) become

$$
\begin{aligned}
L_{i1}^{k1} &= \cdots \delta_{k1} + \cdots, \\
L_{i1}^{k2} &= \cdots \delta_{k2} + \cdots, \\
L_{i1}^{k3} &= \cdots \delta_{k3} + \cdots, \\
L_{i1}^{k4} &= \cdots \delta_{k4} + \cdots, \\
L_{i1}^{k5} &= \cdots \delta_{k5} + \cdots, \\
L_{i1}^{k6} &= \cdots \delta_{k6} + \cdots,
\end{aligned} \tag{73}
$$

$$
\begin{aligned}
L_{i2}^{k1} &= \cdots \delta_{i2} + \delta_{k1} + \cdots, \\
L_{i2}^{k2} &= \cdots \delta_{i2} + \delta_{k2} + N_{i2}^{k2} + \cdots, \\
L_{i2}^{k3} &= \cdots \delta_{i2} + \delta_{k3} + N_{i2}^{k3} + \cdots,
\end{aligned}
$$

$$
\begin{aligned}
L_{i2}^{k4} &= \cdots \delta_{i2} + \delta_{k4} + N_{i2}^{k4} + \cdots, \\
L_{i2}^{k5} &= \cdots \delta_{i2} + \delta_{k5} + N_{i2}^{k5} + \cdots, \\
L_{i2}^{k7} &= \cdots \delta_{i2} + \delta_{k7} + N_{i2}^{k7} + \cdots,
\end{aligned} \tag{74}
$$

$$
\begin{aligned}
L_{i3}^{k2} &= \cdots \delta_{i3} + \delta_{k2} + \cdots, \\
L_{i3}^{k3} &= \cdots \delta_{i3} + \delta_{k3} + N_{i3}^{k3} + \cdots, \\
L_{i3}^{k4} &= \cdots \delta_{i3} + \delta_{k4} + N_{i3}^{k4} + \cdots, \\
L_{i3}^{k5} &= \cdots \delta_{i3} + \delta_{k5} + N_{i3}^{k5} + \cdots, \\
L_{i3}^{k6} &= \cdots \delta_{i3} + \delta_{k6} + N_{i3}^{k6} + \cdots, \\
L_{i3}^{k7} &= \cdots \delta_{i3} + \delta_{k7} + N_{i3}^{k7} + \cdots.
\end{aligned} \tag{75}
$$

Differences can be formed through linear operations. The total operation is a full rank linear transformation, which does not change the least squares solution of the original equations. Single differences can be formed by the following ((73) remains unchanged and therefore will not be listed again):

$$
\begin{aligned}
L_{i2}^{k1} - L_{i1}^{k1} &= \cdots \delta_{i2} + \cdots, \\
L_{i2}^{k2} - L_{i1}^{k2} &= \cdots \delta_{i2} + N_{i2}^{k2} + \cdots, \\
L_{i2}^{k3} - L_{i1}^{k3} &= \cdots \delta_{i2} + N_{i2}^{k3} + \cdots, \\
L_{i2}^{k4} - L_{i1}^{k4} &= \cdots \delta_{i2} + N_{i2}^{k4} + \cdots, \\
L_{i2}^{k5} - L_{i1}^{k5} &= \cdots \delta_{i2} + N_{i2}^{k5} + \cdots, \\
L_{i2}^{k7} &= \cdots \delta_{i2} + \delta_{k7} + N_{i2}^{k7} + \cdots,
\end{aligned} \tag{76}
$$

$$
\begin{aligned}
L_{i3}^{k2} - L_{i2}^{k2} &= \cdots \delta_{i3} - \delta_{i2} - N_{i2}^{k2} + \cdots, \\
L_{i3}^{k3} - L_{i2}^{k3} &= \cdots \delta_{i3} - \delta_{i2} + N_{i3}^{k3} - N_{i2}^{k3} + \cdots, \\
L_{i3}^{k4} - L_{i2}^{k4} &= \cdots \delta_{i3} - \delta_{i2} + N_{i3}^{k4} - N_{i2}^{k4} + \cdots, \\
L_{i3}^{k5} - L_{i2}^{k5} &= \cdots \delta_{i3} - \delta_{i2} + N_{i3}^{k5} - N_{i2}^{k5} + \cdots, \\
L_{i3}^{k6} &= \cdots \delta_{i3} + \delta_{k6} + N_{i3}^{k6} + \cdots, \\
L_{i3}^{k7} - L_{i2}^{k7} &= \cdots \delta_{i3} - \delta_{i2} + N_{i3}^{k7} - N_{i2}^{k7} + \cdots,
\end{aligned} \tag{77}
$$

where two observations are unpaired due to the baseline definitions. Double differences can be formed by

$$
\begin{aligned}
L_{i2}^{k1} - L_{i1}^{k1} &= \cdots \delta_{i2} + \cdots, \\
L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1} &= \cdots N_{i2}^{k2} + \cdots, \\
L_{i2}^{k3} - L_{i1}^{k3} - L_{i2}^{k1} + L_{i1}^{k1} &= \cdots N_{i2}^{k3} + \cdots, \\
L_{i2}^{k4} - L_{i1}^{k4} - L_{i2}^{k1} + L_{i1}^{k1} &= \cdots N_{i2}^{k4} + \cdots, \\
L_{i2}^{k5} - L_{i1}^{k5} - L_{i2}^{k1} + L_{i1}^{k1} &= \cdots N_{i2}^{k5} + \cdots, \\
L_{i2}^{k7} - L_{i2}^{k1} + L_{i1}^{k1} &= \cdots \delta_{k7} + N_{i2}^{k7} + \cdots,
\end{aligned} \tag{78}
$$

$$
\begin{aligned}
L_{i3}^{k2} - L_{i2}^{k2} &= \cdots \delta_{i3} - \delta_{i2} - N_{i2}^{k2} + \cdots, \\
L_{i3}^{k3} - L_{i2}^{k3} - L_{i3}^{k2} + L_{i2}^{k2} &= \cdots N_{i3}^{k3} - N_{i2}^{k3} + N_{i2}^{k2} + \cdots, \\
L_{i3}^{k4} - L_{i2}^{k4} - L_{i3}^{k2} + L_{i2}^{k2} &= \cdots N_{i3}^{k4} - N_{i2}^{k4} + N_{i2}^{k2} + \cdots, \\
L_{i3}^{k5} - L_{i2}^{k5} - L_{i3}^{k2} + L_{i2}^{k2} &= \cdots N_{i3}^{k5} - N_{i2}^{k5} + N_{i2}^{k2} + \cdots, \\
L_{i3}^{k6} &= \cdots \delta_{i3} + \delta_{k6} + N_{i3}^{k6} + \cdots, \\
L_{i3}^{k7} - L_{i2}^{k7} - L_{i3}^{k2} + L_{i2}^{k2} &= \cdots N_{i3}^{k7} - N_{i2}^{k7} + N_{i2}^{k2} + \cdots.
\end{aligned} \tag{79}
$$

Using (78) and (73), (79) can be further modified to

$$
\begin{aligned}
&L_{i3}^{k2} - L_{i2}^{k2} + (L_{i2}^{k1} - L_{i1}^{k1}) + (L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1}) = \cdots \delta_{i3} + \cdots, \\
&L_{i3}^{k3} - L_{i2}^{k3} - L_{i3}^{k2} + L_{i2}^{k2} + (L_{i2}^{k3} - L_{i1}^{k3} - L_{i2}^{k1} + L_{i1}^{k1}) - (L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1}) = \cdots N_{i3}^{k3} + \cdots, \\
&L_{i3}^{k4} - L_{i2}^{k4} - L_{i3}^{k2} + L_{i2}^{k2} + (L_{i2}^{k4} - L_{i1}^{k4} - L_{i2}^{k1} + L_{i1}^{k1}) - (L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1}) = \cdots N_{i3}^{k4} + \cdots, \\
&L_{i3}^{k5} - L_{i2}^{k5} - L_{i3}^{k2} + L_{i2}^{k2} + (L_{i2}^{k5} - L_{i1}^{k5} - L_{i2}^{k1} + L_{i1}^{k1}) - (L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1}) = \cdots N_{i3}^{k5} + \cdots, \\
&L_{i3}^{k6} - L_{i1}^{k6} = \cdots \delta_{i3} + N_{i3}^{k6} + \cdots, \\
&L_{i3}^{k7} - L_{i2}^{k7} - L_{i3}^{k2} + L_{i2}^{k2} + (L_{i2}^{k7} - L_{i2}^{k1} + L_{i1}^{k1}) - (L_{i2}^{k2} - L_{i1}^{k2} - L_{i2}^{k1} + L_{i1}^{k1}) = \cdots - \delta_{k7} + N_{i3}^{k7} + \cdots
\end{aligned} \tag{80}
$$

or

$$
\begin{aligned}
L_{i3}^{k2} - L_{i1}^{k2} &= \cdots \delta_{i3} + \cdots, \\
L_{i3}^{k3} - L_{i1}^{k3} - L_{i3}^{k2} + L_{i1}^{k2} &= \cdots N_{i3}^{k3} + \cdots, \\
L_{i3}^{k4} - L_{i1}^{k4} - L_{i3}^{k2} + L_{i1}^{k2} &= \cdots N_{i3}^{k4} + \cdots, \\
L_{i3}^{k5} - L_{i1}^{k5} - L_{i3}^{k2} + L_{i1}^{k2} &= \cdots N_{i3}^{k5} + \cdots, \\
L_{i3}^{k6} - L_{i1}^{k6} - L_{i3}^{k2} + L_{i1}^{k2} &= \cdots N_{i3}^{k6} + \cdots, \\
L_{i3}^{k7} - L_{i3}^{k2} + L_{i1}^{k2} &= \cdots - \delta_{k7} + N_{i3}^{k7} + \cdots.
\end{aligned} \tag{81}
$$

From the last equations of (78) and (81), it is obvious that the clock error and the ambiguities of satellite $k7$, which is not observed by the reference station, are linearly correlated. Keeping one of the ambiguities of the satellite $k7$ at stations $i2$ or $i3$ is necessary and equivalent. Therefore, for any satellite that is not observed by the

reference station, one of the related ambiguities should be kept fixed (station selection is arbitrary). In other words, one of the ambiguities of all satellites has to be kept fixed. In this way, every transformed equation includes only one bias parameter and the bias parameters are linearly independent (regular). Furthermore, the differencing cannot be formed for the unpaired observations of every baseline. However, in the case of an undifferenced adjustment, the situation would be different. We notice that the equation for $k6$ in (80) can be transformed to a double differencing one in (81). If more data are used in the undifferenced algorithm than in the differencing method, the number of undifferenced ambiguity parameters will be larger than that of the double differencing ones. Therefore, we have to drive the so-called data condition to guarantee that the data are able to be differenced or, equivalently, we have to extend the way of double differencing forming so that the differencing will be not limited by special baseline design. Both will be discussed in Sect. 7.5.2.

The meanings of the parameters are changed by independent parameterisation, and they can be read from (73), (74) and (75). The clock errors of the satellites observed by the reference station include the errors of receiver clock and ambiguities. The receiver clock errors include the error of ambiguity of the reference satellite of the same station. Due to the inseparable property of the bias parameters, the clock error parameters no longer represent pure clock errors, and the ambiguities no longer represent pure physical ambiguity. Theoretically speaking, the synchronisation applications of GPS may not be realised using the carrier-phase observations. Furthermore, (81) shows that the undifferenced ambiguities of $i3$ have the meaning of double differencing ambiguities of the stations $i3$ and $i1$ in this case.

Up to now, we have discussed the correlation problem of the bias parameters and found a method of how to parameterise the GPS observations regularly to avoid the problem of rank defect. Of course, many other ways to parameterise the GPS observation model can be similarly derived. However, the parameter sets should be equivalent to each other and can be transformed from one set to another uniquely as long as the same data are used.

7.4.3 Geometry-Free Illustration

The reason why the reference parameters have to be fixed lies in the nature of range measurements, which cannot provide information of the datum origin (cf., e.g. Wells et al. 1987, p. 9). Suppose d is the direct measurement of clock errors of satellite k and receiver i, i.e. $d_i^k = \delta_i + \delta_k$, no matter how many observations were made and how the indices were changed, one parameter (i.e. reference clock) is inseparable from the others and has to be fixed. Suppose h is the direct measurement of ambiguity N and clock errors of satellite k and receiver i, i.e. $h_i^k = \delta_i + \delta_k + N_i^k$, the number of over-parameterised biases is exactly the number of total observed satellites and used receivers. This ensures again that our parameterisation method to fix the reference clock and one ambiguity of every satellite as well as one ambiguity of the reference satellite of every non-reference station is reasonable. The case of combination of d and h (as code and phase observations) will be discussed in the next section.

7.4.4 Correlation Analysis in the Case of Phase–Code Combinations

A phase–code combined observation equation can be written by (cf. Sect. 7.5.2 of Xu 2007)

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} - \begin{pmatrix} A_{11} & A_{12} \\ A_{11} & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} w_\mathrm{p} P_0 & 0 \\ 0 & w_\mathrm{c} P_0 \end{pmatrix}, \tag{82}$$

where L_1 and L_2 are the observational vectors of phase (scaled in length) and code, respectively; V_1 and V_2 are related residual vectors; X_2 and X_1 are unknown vectors of ambiguity and others; A_{12} and A_{11} are related coefficient matrices; P_0 is a symmetric and definite weight matrix; and w_p and w_c are weight factors of the phase and code observations.

The phase, code and phase–code normal equations can be formed respectively by

$$\begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix},$$

$$N_{11} X_1 = R_\mathrm{c} \quad \text{and}$$

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \tag{83}$$

where

$$\begin{aligned} M_{11} &= (w_\mathrm{p} + w_\mathrm{c}) A_{11}^\mathrm{T} P_0 A_{11} = (w_\mathrm{p} + w_\mathrm{c}) N_{11}, \\ M_{12} &= M_{21}^\mathrm{T} = w_\mathrm{p} A_{11}^\mathrm{T} P_0 A_{12} = w_\mathrm{p} N_{12}, \\ M_{22} &= w_\mathrm{p} A_{12}^\mathrm{T} P_0 A_{12} = w_\mathrm{p} N_{22}, \\ B_1 &= A_{11}^\mathrm{T} P_0 (w_\mathrm{p} L_1 + w_\mathrm{c} L_2) = w_\mathrm{p} R_1 + w_\mathrm{c} R_\mathrm{c} \quad \text{and} \\ B_2 &= w_\mathrm{p} A_{12}^\mathrm{T} P_0 L_1 = w_\mathrm{p} R_2. \end{aligned} \tag{84}$$

The covariance matrix Q is denoted as

$$Q = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \tag{85}$$

where (Gotthardt 1978)

$$\begin{aligned} Q_{11} &= (M_{11} - M_{12} M_{22}^{-1} M_{21})^{-1}, \\ Q_{22} &= (M_{22} - M_{21} M_{11}^{-1} M_{12})^{-1}, \\ Q_{12} &= M_{11}^{-1} (-M_{12} Q_{22}) \quad \text{and} \end{aligned} \tag{86}$$

$$Q_{21} = M_{22}^{-1}(-M_{21}Q_{11}),$$

i.e.

$$\begin{aligned} Q_{11} &= ((w_\mathrm{p} + w_\mathrm{c})N_{11} - w_\mathrm{p}N_{12}N_{22}^{-1}N_{21})^{-1}, \\ Q_{22} &= (w_\mathrm{p}N_{22} - w_\mathrm{p}^2(w_\mathrm{p} + w_\mathrm{c})^{-1}N_{21}N_{11}^{-1}N_{12})^{-1} \quad \text{and} \\ Q_{21} &= -N_{22}^{-1}N_{21}((w_\mathrm{p} + w_\mathrm{c})N_{11} - w_\mathrm{p}N_{12}N_{22}^{-1}N_{21})^{-1}. \end{aligned} \tag{87}$$

Thus the correlation coefficient C_{ij} is a function of w_p and w_c, i.e.

$$C_{ij} = f(w_\mathrm{p}, w_\mathrm{c}), \tag{88}$$

where indices i and j are the indices of unknown parameters in X_1 and X_2. For $w_\mathrm{c} = 0$ (only phase is used, X_1 and X_2 are partly linear correlated) and $w_\mathrm{c} = w_\mathrm{p}$ (X_1 and X_2 are uncorrelated), there exist indices ij, so that

$$C_{ij} = f(w_\mathrm{p},\ w_\mathrm{c} = 0) = 1 \quad \text{and} \quad C_{ij} = f(w_\mathrm{p}, w_\mathrm{c} = w_\mathrm{p}) = 0. \tag{89}$$

In other words, there exist indices i and j, the related unknowns are correlated if $w_\mathrm{c} = 0$ and uncorrelated if $w_\mathrm{c} = w_\mathrm{p}$. In the case of a phase–code combination, $w_\mathrm{c} = 0.01w_\mathrm{p}$ can be selected, and one has

$$C_{ij} = f(w_\mathrm{p},\ w_\mathrm{c} = 0.01w_\mathrm{p}), \tag{90}$$

whose value should be very close to 1 (strong correlated) in the discussed case. Equations (88), (89) and (90) indicate that, for the correlated unknown pair ij, the correlation situation may not change much by combining the code to the phase because of the lower weight of the code related to the phase. A numerical test confirmed this conclusion (Xu 2004).

7.4.5 Conclusions and Comments on Parameterisation

In this section, the singularity problem of the undifferenced GPS data processing is pointed out and an independent parameterisation method is proposed for bias parameters of the GPS observation model. The method is implemented into software, and the results confirm the correctness of the theory and algorithm. Conclusions can be summarised as follows:

1. Bias parameterisation of undifferenced GPS phase observations with all clock errors except the reference one and all undifferenced ambiguities are linearly correlated. The linear equation system of undifferenced GPS is then singular and cannot be solved theoretically.

2. A linear independent bias parameterisation can be reached by fixing the reference clock of the reference station, fixing one of the ambiguities of every satellite of arbitrary station (called reference station of every satellite) and fixing the ambiguities of the reference satellite of every non-reference station. The selections of the references are arbitrary; however, the selections are not allowed to be duplicated.
3. The linear independent ambiguity parameter set is equivalent to the parameter set of double differencing ambiguities, and they can be transformed from one to another uniquely if the same data are used.
4. The physical meanings of the bias parameters are varied depending on the way of parameterisation. Due to the inseparable property of the bias parameters, the synchronisation applications of GPS may not be realised using the carrier-phase observations.
5. The phase–code combination does not change the correlation relation between the correlated biases significantly.

Due to the facts regarding the use of the undifferenced algorithm, it is worth giving some comments:

1. In the undifferenced algorithm, the observation equation is a rank defect one if the over-parameterisation problem has not been taken into account. The numerical inexactness introduced by eliminating the clock error parameters and the use of a priori information of some other parameters are the reasons why the singular problem is solvable in practice so far.
2. Using the undifferenced and differencing methods, solutions of the common parameters must be the same if the undifferenced GPS data modelling is really an equivalent one and not over-parameterised.
3. A singular undifferenced parameterisation may become regular by introducing conditions or by fixing some of the parameters through introducing a priori information.

7.5 Equivalence of the GPS Data Processing Algorithms

The equivalence theorem, an optimal method for forming an independent baseline network and a data condition, and the equivalent algorithms using secondary observables are discussed in this section (cf. Xu et al. 2006c).

7.5.1 Equivalence Theorem of GPS Data Processing Algorithms

In Sect. 7.2 the equivalence properties of uncombined and combining algorithms of GPS data processing are given. Whether uncombined or combining algorithms are used, the results obtained are identical and the precisions of the solutions are

identical, too. It is notable that the parameterisation is very important. The solutions depend on the parameterisation. For convenience, the original GPS observation equation and the solution are listed as (cf. Sect. 7.3)

$$\begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 1 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 1 & -f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix}, P = \begin{pmatrix} \sigma_{\mathrm{c}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\mathrm{c}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\mathrm{p}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\mathrm{p}}^2 \end{pmatrix}^{-1}, \tag{91}$$

and

$$\begin{pmatrix} \lambda_1 N_1 \\ \lambda_2 N_2 \\ B_1 \\ C_\rho \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q & -1/q & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ \lambda_1\Phi_1 \\ \lambda_2\Phi_2 \end{pmatrix}, \tag{92}$$

where the meanings of the symbols are the same as that of (44) and (48).

In Sect. 7.2 the equivalence properties of undifferenced and differencing algorithms of GPS data processing are given. Whether undifferenced or differencing algorithms are used, the results obtained are identical and the precisions of the solutions are equivalent. It is notable that the equivalence here is slightly different from the equivalence in combining algorithms. To distinguish them, we call the equivalence in differencing case a soft equivalence. The soft equivalence is valid under three so-called conditions. The first is a data condition, which guarantees that the data used in undifferenced or differencing algorithms are the same. The data condition will be discussed in the next section. The second is a parameterisation condition, i.e. the parameterisation must be the same. The third is the elimination condition, i.e. the parameter set to be eliminated should be able to be eliminated. (Implicitly, the parameter set of the problem should be a regular one.) Because of the process of elimination, the cofactor matrices of the undifferenced and differencing equations are different. If the cofactor of an undifferenced normal equation has the form of

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} = Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \tag{93}$$

then we call the diagonal part of the cofactor

$$Q_e = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}^{-1} = \begin{pmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{pmatrix} \tag{94}$$

an equivalent cofactor. The equivalent cofactor has the same diagonal element blocks as the original cofactor matrix Q and guarantees that the precision relation between the unknowns remains the same. The soft equivalence is defined as follows: the solutions are identical and the covariance matrices are equivalent. Such a

definition is implicitly used in the traditional block-wise least squares adjustment. It is notable that the parameterisation is very important and the rank of the normal equation of the undifferenced observation equation must be equal to the rank of the normal equation of the differencing observation equation plus the number of the eliminated independent parameters. For convenience, the original GPS observation equation and the equivalent differencing equation can be generally written as (cf. (63) and (66))

$$V = L - \begin{pmatrix} A_1 & A_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad P \tag{95}$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} L \\ L \end{pmatrix} - \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}. \tag{96}$$

In Sect. 7.4 the way to parameterise the GPS observables independently is proposed. A correct and reasonable parameterisation is the key to a correct conclusion by combining and differencing derivations. An example is given in Sect. 7.4.1 where an illogical conclusion is derived due to the inexact parameterisation.

For any GPS survey with a definitive space–time configuration, observed GPS data can be parameterised (or modelled) in a suitable way and listed together in a form of linear equations for processing. Combining and differencing are two linear transformations. Because the uncombined and combining data (or equations) are equivalent, differencing the uncombined or combining equations is (soft) equivalent. Inversely, the combining operator is an invertible transformation; making or not making the combination operation on the equivalent undifferenced or differencing equations (95) and (96) is equivalent. That is, the mixtures of the combining and differencing algorithms are also equivalent to the original undifferenced and uncombined algorithms. The equivalence properties can be summarised in a theorem as follows:

Equivalence Theorem of GPS Data Processing Algorithms
Under the three so-called equivalence conditions and the definition of the so-called soft equivalence, for any GPS survey with definitive space–time configuration, GPS data processing algorithms – uncombined and combining algorithms, undifferenced and differencing algorithms, as well as their mixtures – are at least soft equivalent. That is, the results obtained by using any algorithm or any mixture of the algorithms are identical. The diagonal elements of the covariance matrix are identical. The ratios of the precisions of the solutions are identical. None of the algorithms are preferred in view of the results and precisions. Suitable algorithms or mixtures of the algorithms will be specifically advantageous for special kinds of data dealings.

The implicit condition of this theorem is that the parameterisation must be the same and regular. The parameterisation depends on different configurations of the GPS surveys and strategies of the GPS data processing. The theorem says that if the data used are the same and the model is parameterised identically and regularly, then the results must be identical and the precision should be equivalent. This is a guiding principle for the GPS data processing practice.

7.5.2 Optimal Baseline Network Forming and Data Condition

It is well known that for a network with n stations there are $n-1$ independent baselines. An independent baseline network can be stated in words: all stations are connected through these baselines, and the shortest way from one station to any other station is unique. Generally speaking, a shorter baseline leads to a better common view of the satellites. Therefore, the baseline should be formed so that the length of the baseline falls as short as possible. For a network, an optimal choice should be that the summation of weighted lengths of all independent baselines should be minimal. This is a specific mathematic problem called a minimum spanning tree (cf., e.g. Wang et al. 1979).

Algorithms exist to solve this minimum spanning tree problem with software. Therefore, we will just show an example here. An IGS network with ca. 100 stations and the related optimal and independent baseline tree is shown in Fig. 7.1. The average length of the baselines is ca. 1,300 km. The maximum distance is ca. 3,700 km.

In the traditional double differencing model, the unpaired GPS observations of every designed baseline have to be omitted because of the requirement of differencing. (In the example of Sect. 7.4.2, two observations of $k6$ will be omitted. However, if the differencing is not limited by baseline design, no observations have to be cancelled out.) Therefore, an optimal means of double differencing should be based on an optimal baseline design to form the differencing first, then, without limitation of the baseline design, to check for the unpaired observations in order to form possible differencing. This measure is useful for raising the rate of data used by the differencing method. An example of an IGS network with 47 stations and one day's observations has shown (Xu 2004) that 87.9% of all data are used in difference forming based on the optimal baseline design, whereas 99.1% of all data

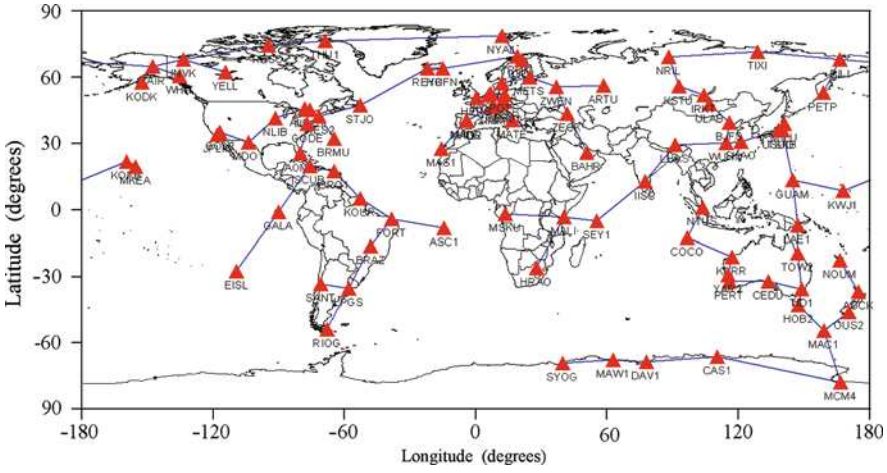


Fig. 7.1 Independent and optimal IGS GPS baseline network (100 stations)

are used in the extended method of difference forming without limitation of the baseline design. That is, the original data may be nearly 100% used for such a means of double differencing.

In the undifferenced model, in order to be able to eliminate the clock error parameters it is sufficient that every satellite is observed at least at two stations (for eliminating the satellite clock errors) and at every station there is a satellite combined with one of the other satellites that are commonly viewed by at least one of the other stations (for eliminating the receiver clock errors). The condition ensures that extended double differencing can be formed from the data. The data have to be cancelled out if the condition is not fulfilled or the related ambiguities have to be kept fixed.

For convenience, we state the data condition as follows:

Data condition: all satellites must be observed at least twice (for forming single differences) and one satellite combined with one of the other satellites should be commonly viewed by at least one of the other stations (for forming double differences).

It is notable that the data condition above is valid for single and double differencing. For triple differencing and user-defined differencing the data condition may be similarly defined. The data condition is one of the conditions of the equivalence of the undifferenced and differencing algorithms. The data condition is derived from the difference forming; however, it is also suggested to use it in undifferenced methods to reduce the singular data. The optimal baseline network forming is beneficial for differencing methods to raise the rate of used data.

7.5.3 Algorithms Using Secondary GPS Observables

As stated in Sects. 7.3 and 7.4, the uncombined and combining algorithms are equivalent. A method of GPS data processing using secondary data is outlined in Sect. 7.3.3. However, a concrete parameterisation of the observation model is only possible after the method of independent parameterisation is discussed in Sect. 7.4. The data processing using secondary observables leads to equivalent results of any combining algorithms. Therefore the concrete parameterisation of the GPS observation model has to be specifically discussed again. The observation model of m satellites viewed at one station is (cf. (67))

$$\begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1\Phi_1(k) \\ \lambda_2\Phi_2(k) \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 1 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 1 & -f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1(k) \\ C_\rho(k) \end{pmatrix}, \; k = 1, \ldots, m, \tag{97}$$

where the relation

$$B_1^z = \frac{1}{m} \sum_{k=1}^{m} B_1(k)/F_k \tag{98}$$

can be used to map the ionospheric parameters in the path directions to the parameter in the zenith direction. The meanings of the symbols are the same as stated in Sect. 7.3. Solutions of (97) are (similar to (68))

$$\begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1(k) \\ C_\rho(k) \end{pmatrix} = \begin{pmatrix} 1-2a & -2b & 1 & 0 \\ -2a & 2a-1 & 0 & 1 \\ 1/q & -1/q & 0 & 0 \\ a & b & 0 & 0 \end{pmatrix} \begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1 \Phi_1(k) \\ \lambda_2 \Phi_2(k) \end{pmatrix}, \; Q(k), k = 1, \ldots, m, \quad (99)$$

where the covariance matrix $Q(k)$ can be obtained by variance–covariance propagation law. The vector on the left-hand side of (99) is called the secondary observation vector. In the case where K satellites are viewed, the traditional combinations of the observation model and the related secondary solutions are the same as (97) and (99), where $m = K$. However, taking the parameterisation method into account, at least one satellite has to be selected as reference and the related ambiguities cannot be modelled. If one were to suppose that the satellite of index K is the reference one, then the first $m = K$–1 observation equations are the same as (97). The satellite K-related observation equations can be written as

$$\begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1 \Phi_1(k) \\ \lambda_2 \Phi_2(k) \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_s^2/f_1^2 & 1 \\ 0 & 0 & f_s^2/f_2^2 & 1 \\ 0 & 0 & -f_s^2/f_1^2 & 1 \\ 0 & 0 & -f_s^2/f_2^2 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1(k) \\ C_\rho(k) \end{pmatrix}, \; k = K, \quad (100)$$

where the ambiguities are not modelled and the constant effects will be absorbed by the clock parameters. Solutions of (100) are

$$\begin{pmatrix} \lambda_1 N_1(k) \\ \lambda_2 N_2(k) \\ B_1(k) \\ C_\rho(k) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/q & -1/q & -1/q & 1/q \\ 1/2 & 1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} R_1(k) \\ R_2(k) \\ \lambda_1 \Phi_1(k) \\ \lambda_2 \Phi_2(k) \end{pmatrix}, \; Q(K). \quad (101)$$

It is notable that the solutions of the traditional combinations are (99) with $m = K$, whereas for the combinations with independent bias parameterisation, the solutions are the combinations of (99) with $m = K$–1 and (101). It is obvious that the two solutions are different. Because the traditional observation model used is an inexact one, the solutions of the traditional combinations are also inexact. The bias effects (of ambiguity) that are not modelled are merged into the clock bias parameters. Due to the fact that the bias effects cannot be absorbed into the non-bias parameters, only the clock error parameters will be different in the results and the clock errors will have different meanings. Further, the ionosphere-free combinations and geometry-free combinations are correct under the independent parameterisation.

It shows that through exact parameterisation, the combinations are no longer independent from satellite to satellite. For surveys with multiple stations, through

correct parameterisation the combinations will no longer be independent from station to station. Therefore, traditional combinations will lead to incorrect results because of the inexact parameterisation.

The so-called secondary observables on the left-hand side of (99) and (101) can be further processed. The original observables can be uniquely transformed to secondary observables. The secondary observables are equivalent and direct measurements of the ambiguities and ionosphere as well as geometry. Any further GPS data processing can be based on the secondary observables (cf. Sect. 7.3).

7.5.4 Non-equivalent Algorithms

As stated in the equivalence theorem of GPS algorithms, the equivalence properties are valid for GPS surveys with definitive space–time configuration. As long as the measures are the same and the parameterisation is identical and regular, the GPS data processing algorithms are equivalent. It is notable that if the surveys and the parameterisation are different, then the algorithms are not equivalent to each other. For example, algorithms of single-point positioning and multi-points positioning, algorithms of orbit-fixed positioning and orbit co-determined positioning, algorithms of static and kinematic as well as dynamic applications, etc. are non-equivalent algorithms.

7.6 Inferences of Equivalence Principle

7.6.1 Diagonalisation Algorithm

In the above discussion, one group of unknowns is eliminated by matrix partitioning to obtain an equivalently eliminated normal equation system of the other group of unknowns. Using the elimination process twice for the two groups of unknowns, the normal equation can be diagonalised. The algorithm can be outlined as follows.

A linearised observation equation and the normal equations can be represented by (1) and (2). From (3), one has

$$X_1 = M_{11}^{-1}(B_1 - M_{12}X_2). \tag{102}$$

Setting X_1 into (4), one gets an equivalently eliminated normal equation of X_2:

$$M_2 X_2 = R_2, \tag{103}$$

where

$$\begin{aligned} M_2 &= M_{22} - M_{21}M_{11}^{-1}M_{12}, \\ R_2 &= B_2 - M_{21}M_{11}^{-1}B_1. \end{aligned} \tag{104}$$

Similarly, from (4), one has

$$X_2 = M_{22}^{-1}(B_2 - M_{21}X_1). \tag{105}$$

Setting X_2 into (3), one gets an equivalently eliminated normal equation of X_1:

$$M_1 X_1 = R_1, \tag{106}$$

where

$$\begin{aligned} M_1 &= M_{11} - M_{12} M_{22}^{-1} M_{21}, \\ R_1 &= B_1 - M_{12} M_{22}^{-1} B_2. \end{aligned} \tag{107}$$

Combining (106) and (103) together, one has

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, \tag{108}$$

where (cf., e.g. Xu 2007, p. 143)

$$\begin{aligned} Q_{11} &= M_1^{-1}, & Q_{22} &= M_2^{-1}, \\ Q_{12} &= -M_{11}^{-1}(M_{12} Q_{22}), & Q_{21} &= -M_{22}^{-1}(M_{21} Q_{11}). \end{aligned} \tag{109}$$

It is obvious that (2) and (108) are two equivalent normal equations. The solutions of the both equations are identical. Equation (108) is a diagonalised normal equation related to X_1 and X_2. The process of forming (108) from (2) is called the diagonalisation process of a normal equation.

As discussed in Sect. 7.2, the equivalently eliminated observation equation of the second equation of (108) is (9). Similarly, the equivalently eliminated observation equation of the first equation of (108) is (11). Denote

$$D_1 = (E - K) A \quad \text{and} \tag{110}$$

$$D_2 = (E - J) B \tag{111}$$

in (9) and (11), then write them together

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} L \\ L \end{pmatrix} - \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}. \tag{112}$$

Equation (112) is derived from the normal equation (108); therefore, it is true inversely, i.e. (108) is the least squares normal equation of the observation equation (112). So (112) is an equivalent observation equation of (1). Equations (108) and (112) are called diagonalised equations of (2) and (1), respectively.

In the sequential adjustment and Kalman filtering, there is a problem which has not been described exactly and theoretically in the literature before, that is, how to give up the nuisance parameter-related information from the past. For a real-time data processing, it is especially important to keep the updated problem as small as possible. This so-called diagonalisation algorithm is proposed initially for deriving an equivalent ambiguity search criterion and then used for giving up the nuisance parameter-related information (Xu 2003a).

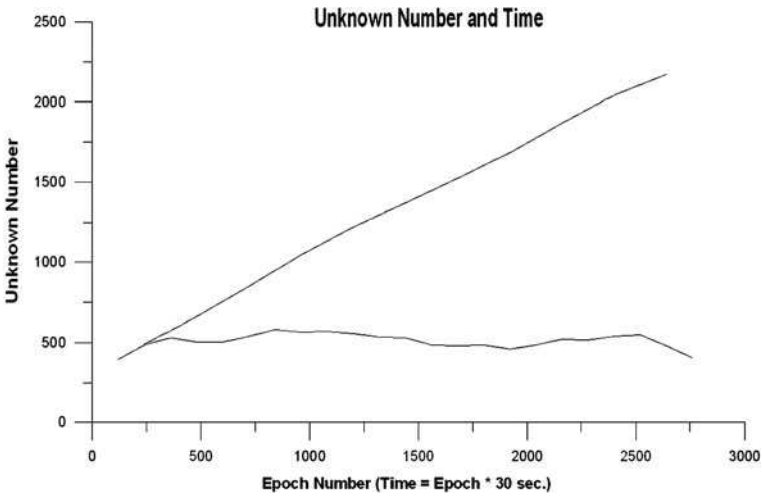


Fig. 7.2 Unknown numbers in non-diagonalised and diagonalised algorithms

Suppose the past surveying information is presented in a normal equation system

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \tag{113}$$

and the normal equation of the present epoch block consists of only the parameter sub-vector X_2, then (113) can be diagonalised by

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \tag{114}$$

so that one just needs to accumulate the X_2 related part in (114) into the present normal equation and then solve the problem. In such a way X_1 is considered as a nuisance parameter sub-vector and is given up during the sequential data processing so that the data processing problem could be kept as small as possible.

Figure 7.2 shows the relationship between unknown number and time for a sequential solution of IGS network with 47 stations and fixed orbit for a network monitoring problem. Without the diagonalisation algorithm, the parameter number increases to 2,300 whereas with the diagonalisation algorithm, the parameter number remains around 500. Without such an algorithm an exact and effective real-time data sequential processing and filtering is not realistic.

7.6.2 Separability of the Observation Equation

For any observation equation (1), instead of forming the least squares normal equation (2), one may form the equivalent observation sub-equations of (112). Then the normal equations of (112) are (108). The total equation (1) is separated into two sub-equations. If the observation equation (1) is not solvable due to the huge number of

unknowns and the capacity limit of the computer, after a suitable separation, the sub-equations may become solvable. To form the two observation sub-equations is more complicated than to form (1) straightforward. However, an unsolvable problem is turned out to be solvable using the proposed separation algorithm.

In practice, the normal equation is first formed from every observation equation and then accumulated. In the separation algorithm, however, the observation equations first have to be formed until the equations can be separated (eliminated), and then the normal equation can be formed and accumulated. The processes of equation-forming and eliminating and then normal equation-forming and accumulating have to be repeated until the complete equations are formed.

That is, any observation equation and its normal equation are separable by using linear algebra operations. Any observation equation can be separated into two sub-equations. The original observation equation may not be solvable sometimes due to the huge number of unknowns and the capacity limit of the computer. However, after a suitable separation, the sub-equations may become solvable.

7.6.3 Optimal Ambiguity Search Criteria

The ambiguity search criterion has been a controversial topic in international GPS research community for a few years.

Suppose GPS observation equation is (1) and its least squares normal equation is (2), where $X_2 = N$ (N is the ambiguity sub-vector) and $X_1 = Y$ (Y is the rest unknown sub-vector). The least squares ambiguity search (LSAS) criterion (Teunissen 1995; Leick 2004; Hofmann-Wellenhof et al. 2001; Euler and Landau 1992; Han and Rizos 1997) is

$$\delta(dN) = (N_0 - N)\mathrm{Inv}(Q_{22})(N_0 - N), \tag{115}$$

where N_0 is the float solution of the ambiguity sub-vector, $dN = N_0 - N$. The ambiguity search is a process to find out a vector N in the searching area so that the value of $\delta(dN)$ reaches the minimum.

The so-called general ambiguity search criterion is derived in Xu (2002b) and has a form of

$$\delta(dX) = (X_0 - X)^{\mathrm{T}}\mathrm{Inv}(Q)(X_0 - X), \tag{116}$$

where $X = (Y\ N)^{\mathrm{T}}$, $X_0 = (Y_0\ N_0)^{\mathrm{T}}$, $dX = X_0 - X$ and index 0 denotes the float solution. The search is a process to find out a vector X (includes N in the searching area and Y computed) so that the value of $\delta(dX)$ reaches the minimum. The optimal property of this criterion can be found in Xu (2002b).

For the equivalent (or diagonalised) normal equation (114), the related equivalent criterion is

$$\begin{aligned}\delta_1(dX) &= (\mathrm{Y}_0 - Y)^{\mathrm{T}}\mathrm{Inv}(Q_{11})(Y_0 - Y) + (N_0 - N)\mathrm{Inv}(Q_{22})(N_0 - N),\\ &= \delta(dY) + \delta(dN)\end{aligned} \tag{117}$$

where index 1 is used to distinguish the equivalent criterion from the general one given in (116). It is obvious that the equivalent criterion is not an exact one theoretically. Expanding (116) and comparing the results with (117), one may notice that two correlation terms are missing (Cai et al. 2007). Numerically, the omission of the two terms may possibly be allowed (Morujao and Mendes 2007).

It is worth mentioning that the well-known ambiguity function (AF) method is mathematically incorrect (cf. Xu 2003b, pp. 280–283) and the so-called least squares (LS) criterion is generally not optimal (cf. Xu 2003b, pp. 153–172). For information to the readers, a brief discussion is given below concerning the LS criterion.

An excellent summary of the derivation of the ambiguity search LS criterion is given in Verhagen (2004). In terminology of Teunissen (1995) the observation model is

$$y = A\,a + B\,b + e, \qquad Q_y \tag{118}$$

and the solution of (118) can be obtained by the following minimization problem:

$$\min_{a,b} \|y - Aa - Bb\|^2_{Q_y}, \quad a \in Z^n,\ b \in R^n. \tag{119}$$

Equation (119) can be orthogonalised by

$$\|y - Aa - Bb\|^2_{Q_y} = \left\|\hat{e}\right\|^2_{Q_y} + \left\|\hat{a} - a\right\|^2_{Q_y} + \left\|\hat{b}(a) - b\right\|^2_{Q_y}. \tag{120}$$

The explanations of the symbols can be found in Verhagen (2004). The solution of (119) is then obtained in three steps: (1) compute the float solution of (118); (2) minimise the second term (denoted by δ_a for later convenience) on the right-hand side of (120) and set the third term (denoted by δ_b) to zero to get the fixed ambiguity; (3) compute the fixed coordinates using fixed ambiguity.

The key problem is that δ_b is not allowed to be set to zero. The significant opposite arguments are listed below (Xu 2004):

- To set δ_b to zero or any constant is equivalent; however, the δ_b is neither zero nor constant for all coordinate vectors computed from all to-be-searched ambiguity candidate vectors. Therefore to set δ_b to zero is indeed to omit the term δ_b in (120). (In other words, in the second step the effect of the term δ_b has never been taken into account in this way.)
- To omit the term δ_b in the minimisation problem (120) is allowed only if the following assumption is true, i.e. a minimum δ_a will lead to a minimum δ_b and therefore lead to a minimum of $(\delta_a + \delta_b)$. However, theory and numerical examples have shown that such an assumption cannot be generally true (cf. Xu 2002b). The omission of the term δ_b in (120) destroys the equivalence between (119) and (120). The problem of minimisation of $(\delta_a + \delta_b)$ is degraded to a problem of minimisation of δ_a. In this way the obtained solution cannot be generally the same as the solution obtained directly from (119).

- Noting optimality und uniqueness properties of the solution of (118), the solution obtained by minimising δ_a cannot generally be the optimal ones of (118). Any indication obtained by such non-optimal results may not really be true.
- To set δ_b to zero is the same as to set b as float solution (then δ_b is zero). However, b cannot be the float solution. Proof: according to the summarised method select $b =$ float coordinate solution; through the minimisation of δ_a one gets the vector a, then one can compute b using a; because a is not the float ambiguity, b is not the float coordinate solution. So the result of b states that b does not equal float coordinate solution; this is in conflict with the assumed starting value.
- To set δ_b to a constant is not allowed either. Proof: for $b =$ float coordinate solution + any constant vector (e.g. $b = 0$ or $b =$ float solution + constant vector of 1,000 km) one has $\delta_b =$ constant, then following $\text{mini}(\delta_a + \delta_b) = \text{mini}(\delta_a)$; the sought for results are the same by using the summarised method, no matter the computation is started from given float position or 1,000 km farther away. Therefore setting $\delta_b =$ constant would be incorrect.

7.7 Summary

Differential and un-differential GPS algorithms, combined and un-combined GPS algorithms as well as their mixtures are equivalent. As soon as the GPS data are measured, the information contents of the data are definitive ones. If the models used are the same and the principle of the adjustment and filtering are also the same, the results obtained should be equivalent. Advantages and disadvantages are relative and balanced. From the equivalence principle, a very important topic of independent parameterisation of the GPS observation model is discussed which points out where the singularity problem comes from. The inferences of the equivalence principle are important beyond the principle itself. Diagonalisation algorithm could be important even for classic adjustment for reducing parameters. Separability of any observation equation and its normal equation may lead to a non-solvable problem becoming a solvable one. The optimal criterion for ambiguity search may clear a confusion caused by the so-called LSSA method. More studies of the equivalence principle are available (Shen and Xu 2007; Shen et al. 2008).

Acknowledgements This study is under the framework of the internal GPS/Galileo study program of GFZ Potsdam and is partly supported by the Overseas Outstanding Chinese Scholar Fund of the Chinese Academy of Sciences hosted by the Institute of Geodesy and Geophysics in Wuhan in cooperation with co-authors.

References

Blewitt, G. (1998) GPS data processing methodology. In: Teunissen, P.J.G. and Kleusberg, A. (eds) GPS for Geodesy. Springer, Berlin, Heidelberg, New York, pp. 231–270

Cai, J., Hu, C. and Grafarend, E.W. (2007) The optimal regularization method and its application in GNSS Rapid Static Positioning, ION GNSS Proceedings, Fort Worth, pp. 299–305

Cannon, M.E., Lachapelle, G., Szarmes, M., Herbert, J., Keith, J. and Jokerst, S. (1997) DGPS kinematic carrier phase signal simulation analysis for precise velocity and position determination. Proceedings of ION NTM 97, Santa Monica

Euler, H.J. and Landau, H. (1992) Fast GPS ambiguity resolution on-the-fly for real-time applications. Proceedings of 6th International Geodetic Symposium on Satellite Positioning, Columbus, pp. 17–20

Gotthardt, E (1978) Einführung in die Ausgleichungsrechnung. Herbert Wichmann Verlag, Karlsruhe

Han, S. and Rizos, C. (1997) Comparing GPS Ambiguity Resolution Techniques. GPS World, 8(10), 54–61

Hofmann-Wellenhof, B., Lichtenegger, H. and Collins, J. (2001) Global Positioning System: Theory and Practice, 5th Ed. Springer, Berlin, Heidelberg, New York

Hugentobler, U., Schaer, S. and Fridez, P. (2001) Bernese GPS Software: Version 4.2. Astronomical Institute University of Berne, Berne

ION-proceedings: Proceedings of ION GPS-1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008

Jong, P.J. de (1998) A processing strategy of the application of the GPS in network, PhD thesis, Netherlands Geodetic Commission, Delft

Kouba, J. and Heroux, P. (2001) Precise point positioning using IGS orbit and clock products. GPS Solut., 5(2), 12–28

Langley, R.B. (1998a) Propagation of the GPS signals. In: Teunissen, P.J.G. and Kleusberg, A. (eds) GPS for Geodesy. Springer, Berlin, Heidelberg, New York, pp. 111–149

Langley, R.B. (1998b) GPS receivers and the observables. In: Teunissen, P.J.G. and Kleusberg, A. (eds) GPS for Geodesy. Springer, Berlin, Heidelberg, New York, pp. 151–185

Leick, A. (2004) GPS Satellite Surveying, 3rd Ed. Wiley, New York

Morujao, D.B. and Mendes, V.B. (2007) Investigation of Instantaneous Carrier Phase Ambiguity Resolution with the GPS/Galileo Combination using the General Ambiguity Search Criterion. J. GPS, 7(1), 35–45

Parkinson, B.W. and Spilker, J.J. (eds) (1996) Global Positioning System: Theory and Applications. Vols. I, II, Progress in Astronautics and Aeronautics, Vol. 163. American Institute of Aeronautics and Astronautics, Inc., Washington

Petovello, M.G. (2006) Narrowlane: is it worth it? GPS Solut., 10(3), 187–195

Remondi, B. (1984) Using the Global Positioning System (GPS) phase observable for relative geodesy: modelling, processing, and results. University of Texas at Austin, Center for Space Research, Austin

Schafrin, B. and Grafarend, E. (1986) Generating classes of equivalent linear models by nuisance parameter elimination, Applications to GPS observations. Manuscripta Geodetica 11, 262–271

Shen, Y.Z., Li, B.F., Xu G (2008) Supplement to "Simplified equivalent representation of GPS observation equations". GPS Solut., 12, 99–108

Shen, Y.Z. and Xu, G. (2007) Simplified equivalent representation of differenced GPS observation equations. GPS Solut., doi: 10.2007/s10291-007-0070-z

Sjoeberg, L.E. (1999) Unbiased vs. biased estimation of GPS phase ambiguities from dual-frequency code and phase observables. J. Geodesy, 73, 118–124

Strang, G. and Borre, K. (1997) Linear Algebra, Geodesy and GPS. Wellesley-Cambridge Press, Wellesley

Teunissen, P.J.G. (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geodesy, 70(1–2), 65–82

Verhagen, S. (2004) Integer ambiguity validation: an open problem? GPS Solut., 8(1), 36–43

Wang, L.X., Fang, Z.D., Zhang, M.Y., Lin, G.B., Gu, L.K., Zhong, T.D., Yang, X.A., She, D.P., Luo, Z.H., Xiao, B.Q., Chai, H. and Lin, D.X. (1979) Mathematic handbook. Educational Press, Peking, ISBN 13012-0165

Wells, D., Lindlohr, W., Schaffrin, B. and Grafarend, E. (1987) GPS Design: Undifferenced Carrier Beat Phase Observations and the Fundamental Differencing Theorem. University of New Brunswick, New Brunswick

Wu, S.-C. and Melbourne, W.G. (1993) An optimal GPS data processing technique for precise positioning. IEEE Trans. Geosci. Remote Sens., 31, 146–152, January 1993

Xu, G. (2002a) GPS data processing with equivalent observation equations. GPS Solut., 6(1–2), 28–33

Xu, G. (2002b) A general criterion of integer ambiguity search. J. GPS, 1(2), 122–131

Xu, G. (2003a) A diagonalization algorithm and its application in ambiguity search. J. GPS, 2(1), 35–41

Xu, G. (2003b, 2007) GPS – Theory, Algorithms and Applications. Springer, Heidelberg, Berlin

Xu, G. (2004) MFGsoft – Multi-Functional GPS/(Galileo) Software – Software User Manual, (Version of 2004), Scientific Technical Report STR04/17 of GFZ Potsdam, www.gfz-potsdam.de/bib/pub/str0417/0417.pdf

Xu, G., Guo, J. and Yeh, T.K. (2006a) Equivalence of the uncombined and combining GPS algorithms, published in Xu 2007 as Sect. 6.7

Xu, G., Sun, H. and Shen, Y. (2006b) On the parameterisation of GPS observation model, published in Xu 2007 as Sect. 9.1

Xu, G., Yang, Y.X. and Zahng, Q (2006c) Equivalence of the GPS Data Processing Algorithms, published in Xu 2007 as Sect. 9.2

Zhou, J (1985) On the Jie factor. Acta Geodaetica et Geophysica, 5 (in Chinese)

Zumberge, J.F., Heflin, M.B., Jefferson, D.C., Watkins, M.M. and Webb, F.H. (1997) Precise point positioning for the efficient and robust analysis of GPS data from large networks. J. Geophys. Res., 102(B3), 5005–5017

Chapter 8
Marine Geodesy

Joerg Reinking

Contents

8.1 Introduction

Taking a look at our planet, we can easily realize that more than 70% of the Earth's surface is covered by oceans and coastal waters. These waters have been the driving force of the climatic changes during the history of the planet. It is also the place where life started to develop and still the largest biosphere.

Oceans and coastal waters are not only important from the ecological point of view but from the aspect of economics. Many gas and oil fields have been discovered and exploited during the last century and the role of methane hydrate, which can be found on the floor of almost all oceans, for industrial use has not even been identified. The globally increased demand for all kinds of seafood has created a large and still growing fishing industry which will satisfy their claims in national and international waters. Furthermore, more than 90% of all goods that are exchanged

J. Reinking (✉)
Jade University of Applied Sciences, Wilhelmshaven/Oldenburg/Elsfleth, Germany
e-mail: reinking@jade-hs.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_8,

on the global market are nowadays transported in seagoing ships that are getting bigger and bigger.

These few simple facts should be sufficient for geodesists to include the hydrosphere in their scope of work. Actually there are many scientists and engineers of our branch working in this field but most of them do not know that all their activities can be summarized under the term “marine geodesy”.

Marine geodesy covers a very wide range of work that is related to oceans and coastal waters. It is hard to define at least some main topics from the large variety of activities without ignoring some important areas. To get an overview on what geodesists are working on, this chapter will focus only on two main topics – bathymetry and hydrography and precise navigation.

8.2 Bathymetry and Hydrography

In bathymetry and hydrography, the object under investigation is generally the seabed. By use of modern hydrographic surveying methods the depth is determined from hydroacoustic observations. In combination with determinations of the horizontal and vertical position of a survey vessel, the seafloor topography and morphology is derived.

The depth of the seafloor is described in relation to a defined reference surface. The definition of this surface is strongly linked to the application of the survey results. If, for example, the hydroacoustic data are used for scientific investigations on the geological seafloor structure, the reference surface can be chosen with the only limitation being steadiness in time. The focus is mainly put on the information about depths in relation to the identical surface and not on an exact vertical datum definition. Those observations are often carried out in deep sea areas and they are typically called bathymetric measurements.

On the other hand, if the results should be used, for example, to guide a ship safely through an approach channel with tidal influence, the reference surface must be connected to the lowest possible water level. The definition of the vertical datum is strongly limited by the users’ needs. Depth observation with a strongly restricted definition of a vertical datum should be referred to as hydrographic measurements. That means that hydrography also has to include the estimation of tides, currents and waves.

8.2.1 Scope of Work

While bathymetric results attempt to give the best representations of the actual seabed that can be used for scientific and other purposes, hydrographic products tend to overrepresent least depths and ignore the actual submarine topography because they are often used as the mariner’s tools to avoid accidents. Although there are still some differences in their outcome, the working areas of bathymetry and hydrography should be considered as a unit because of the use of very similar observation and analysis methods.

The work area of bathymetry and hydrography encloses many different activities and the description of them would go beyond the scope of this chapter. Therefore only some important and interesting aspects will be presented in this context.

8.2.1.1 Echo Soundings of Oceans and Coastal Waters

The observation of water depth is the most important task in bathymetry and hydrography. The depth is nowadays observed by hydroacoustic echo sounders that are described later on. The transceiver of an echo sounder is commonly installed in a ship's hull near the keel.

The distance between the echo sounder and the bottom of the sea is measured by the time of travel of an acoustic signal. To derive the water depth related to a vertical datum this measurement must be corrected for the actual height of the sounder. If a precise GPS system is used on board the ship, the soundings could be related to a reference height system by the resulting GPS height and the distance between the sounder and the GPS antenna (Fig. 8.1a).

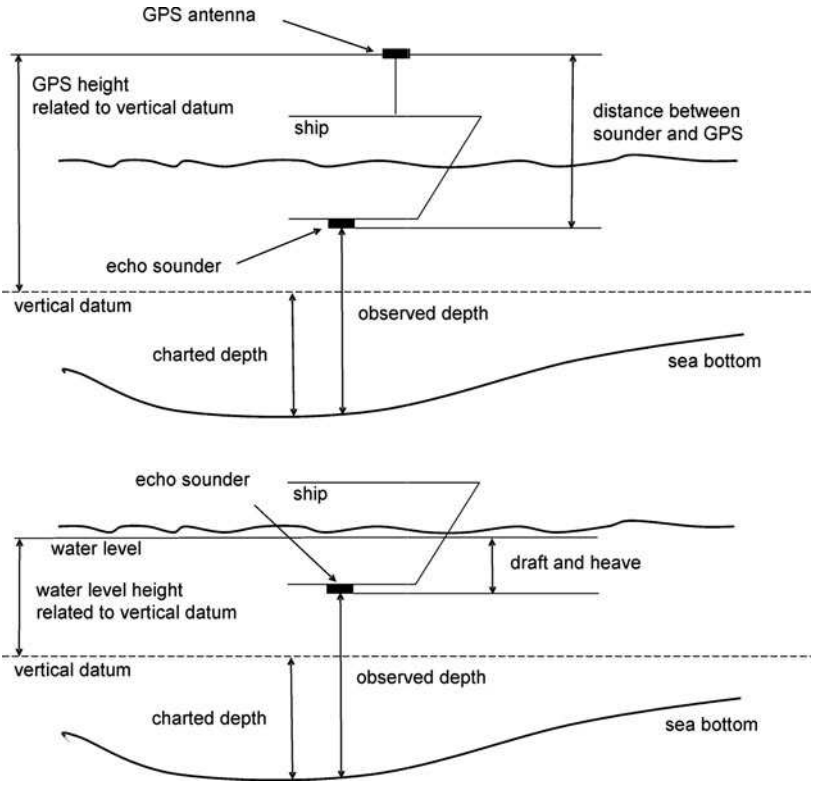


Fig. 8.1 Geometry of hydrographic survey (**a**) (*top*) related to GPS heights and (**b**) (*bottom*) related to actual water level

If no precise GPS height is available, the sounding has to be related to the actual water level that must be derived from tide gauge readings and oceanographic models of the sea surface. Additionally the static and dynamic draught and the wave- and swell-induced vertical movement of the ship (described as heave) have to be observed and taken into account (Fig. 8.1b).

The definition of the vertical datum depends on the intended purpose of the resulting data. Bathymetric observations are often simply related to an ellipsoid and the depths are presented as ellipsoidal heights. This is sufficient for a detailed description of the seabed, i.e. for scientific analysis. Hydrographic data are frequently used as navigational information and mariners are interested in knowledge of the minimum water depth for a safe passage. Therefore the vertical datum is commonly connected to the lowest possible water level and defined in most nations as the lowest astronomical tide (LAT).

8.2.1.2 Seafloor Maps

For the creation of seafloor maps, depths from echo soundings have to be related to the horizontal position of the measurement. Fortunately, the quality demands for the positions are not as strict as for geodetic surveys on land. Concerning hydrographic surveys, the accuracy standards are defined by the International Hydrographic Organization (IHO) as 2–5 m for harbours or coastal areas and 20–150 m in offshore areas (95% confidence level) (IHO 1998). This accuracy can be achieved sufficiently with DGPS or GPS code solutions.

The quality of the map depends strongly on the needs of the map user and the purpose of the mapping. Common bathymetric maps show a detailed structure of the seafloor which could be used for scientific investigation and boundary demarcation (Fig. 8.2a). Hydrographic maps are widely used in navigation and therefore the seafloor information is generalized to the essential depth data and, particularly in open waters, only some important least depths are displayed (Fig. 8.2b).

8.2.1.3 Scientific Investigations

Particularly bathymetric data are often used in scientific investigation related to geology, geophysics and glaciology. Precise results from deep sea echo sounding give information on seafloor faults and seamount structures as well as on the behaviour of glaciers in past times and the morphology of subduction zones.

The last research cruise of the German research vessel *SONNE*, for example, was used to carry out bathymetric observations at the continental margin of Sumatra (Fig. 8.3). The resulting data helped to answer geological questions concerning possible canyons crossing the continental shelf of Sumatra that are probably created by underwater rivers. Additionally these data are to be used for the positioning of GPS tsunami buoys.

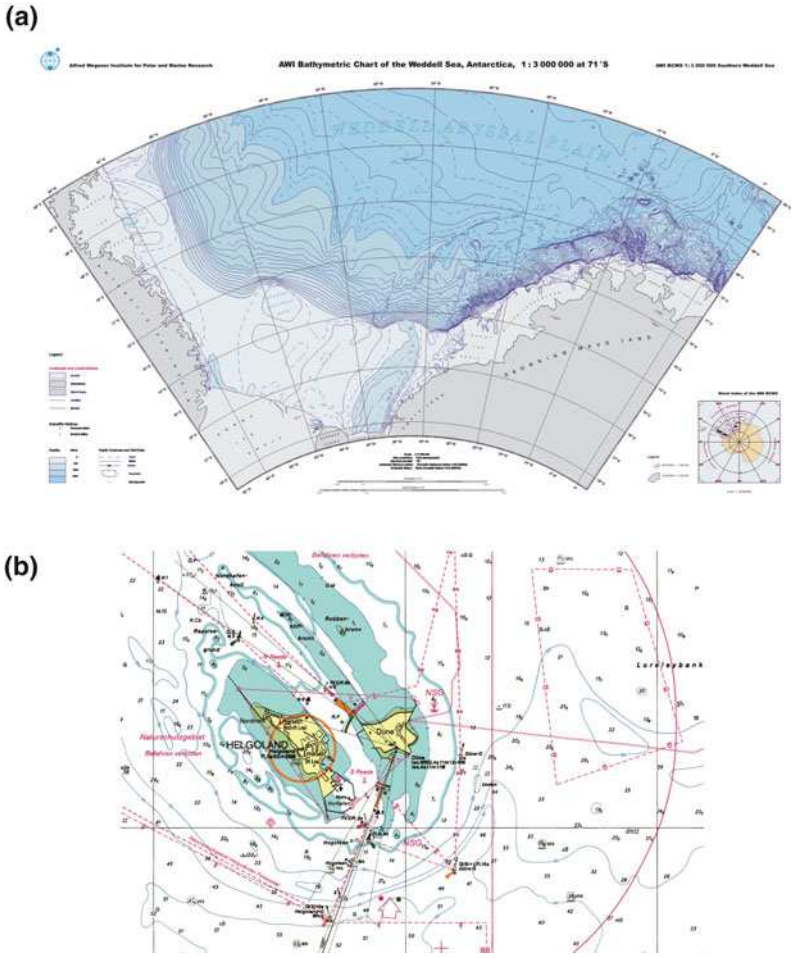


Fig. 8.2 (**a**) Bathymetric map (Schenke et al. 1998) and (**b**) hydrographic map (BSH Chart 3 Ansteuerung Helgoland, 2004)

8.2.1.4 Boundary Demarcation and Determination

The marine boundaries of nations are defined by the United Nations Convention on the Law of the Sea (UNCLOS). The sovereignty of a coastal state extends up to 12 nautical miles from the baseline of the state. The baseline of the territorial sea is the low waterline along the coast which has to be observed by hydrographic methods (UNCLOS 2006).

An exclusive economic zone (EEZ) is an area beyond and adjacent to the territorial sea. The coastal state has sovereign rights for the purpose of exploring and exploiting, conserving and managing the natural resources. The EEZ shall not extend beyond 200 nautical miles from the baselines of the territorial sea. States with overlapping EEZ have to define the limit of the EEZ by mutual agreements.

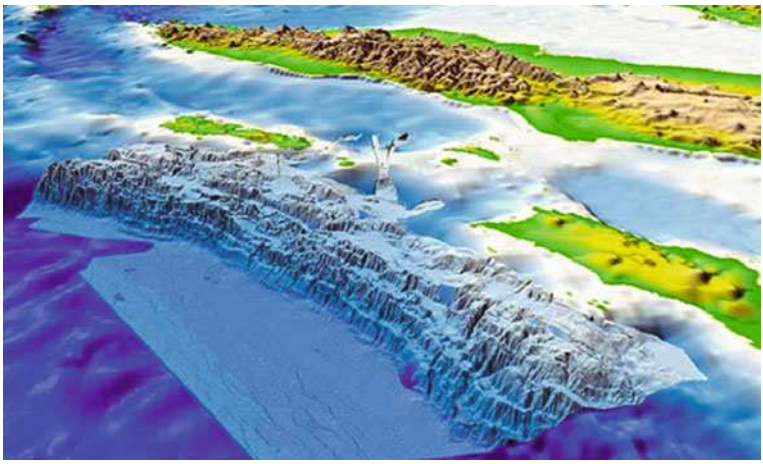


Fig. 8.3 Three-dimensional model from bathymetric survey of Sumatra (Ladage et al. 2006)

The economic influence area therefore strongly depends on the results of hydrography. The importance of marine geodesy is easily realized if we take a look at, e.g., Australia (Fig. 8.4a). The total land area of this state is approximately 7.7 million km^2 but the EEZ with an area of 8.1 million km^2 is larger than the whole "country".

Particularly the rights of exploitation and exploration are of great economic interest. According to UNCLOS, states can extend their claims up to the limit of the continental shelf (CS). This limit is defined by different slope, distance and sediment thickness criteria and ends at least 350 nm from the baseline or at a depth of 2,500 m. All those definitions are only applicable if sufficient hydrographic and bathymetric data are available. In the case of Australia, the CS adds some 2 million km^2 to its zone of economic interest (Fig. 8.4b).

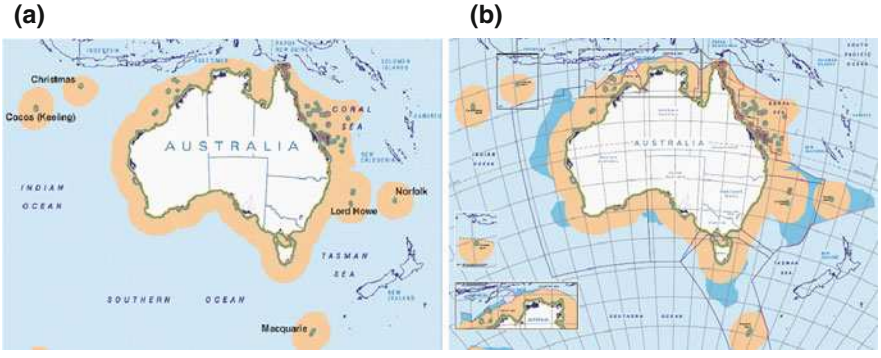


Fig. 8.4 (**a**) Australia and its EEZ and (**b**) Australia and its continental shelf. © Geoscience Australia (2006). All rights reserved

8.2.2 Hydroacoustic Measurements

Hydroacoustic measurement is the most important observation technique in bathymetry and hydrography. It is not only used for the determination of water depths but also in current profiling and underwater positioning.

At the beginning of the twentieth century, the German physicist Dr. Alexander Behm was one of the pioneers in underwater sound propagation. After the accident of the *TITANIC* in April 1912 he worked on an iceberg warning system based on the sound transmission from a ship to the sea. Unfortunately the echoes from icebergs were very poor and the system did not work well in iceberg detection but he recognized strong reflection from the seafloor. So he invented the echo sounder more or less by chance and patented it in 1913. Although the technique of echo sounders developed rapidly during the last 100 years, the basic principles of hydroacoustic measurements are still valid.

8.2.2.1 Basic Principles

Hydroacoustic depth measurements are based on the observation of travel time of acoustic waves in water. An acoustic pulse is generated by a transmitter and forwarded to a receiver/transmitter (T/R) switch. The T/R switch passes the power to a transducer commonly mounted in a ship's hull. The transducer converts the electrical power into acoustic power and sends it in the form of an acoustic signal into the water. The signal propagates beam-shaped in the water with an opening angle depending on the emitted signal frequency and the transducer's diameter. It is reflected by the seafloor and the echo is received at the transducer which converts it into an electrical signal. The signal is sent through the T/R switch to the receiver which amplifies the signal and sends it to the recorder and display unit. The recorder controls the signal emission, measures the travel time of the acoustic signal, stores the data and converts time intervals into ranges (Fig. 8.5) (according to De Jong et al. 2002).

A quartz clock inside the transmitter is used to obtain the operating frequency of the transducer and to measure the time intervals between the signal transmission and reception. The depth is simply calculated from

$$d = \frac{\Delta t}{2} \times c_w$$

where Δt is the two-way travel time and c_w is the speed of sound in the water. The speed of sound mainly depends on the salinity and the temperature of the water and in greater depth additionally on the pressure. Different empirical formulae have been developed to calculate the speed of sound, depending on parameters mentioned above. The speed of sound reaches values of 1,450 m/s in a depth of 500 m in fresh water with a temperature of 10°C and 1,495 m/s for ocean water with a salinity of 3% under the same depth and temperature conditions. For larger depth it is strongly recommended to use a sound speed profile that varies according

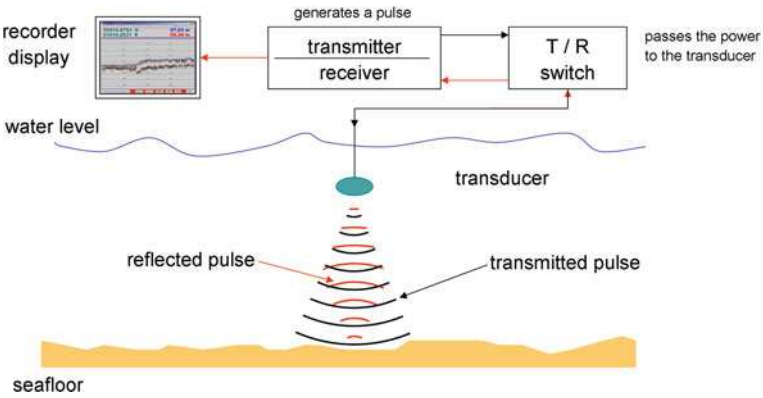


Fig. 8.5 Basic principle of hydroacoustic depth measurement

to the depth. A very efficient way to produce a current sound speed profile is to use a sound velocity probe which measures the travel time of a sound over a fixed distance. Standard instruments are able to operate autonomously so they can be deployed to different water depths on a line and observe a complete sound velocity profile.

A very important basic fact in hydroacoustics is that the reflection of a signal depends on the frequency of the emitted sound signal as well as on the density of the reflecting bottom's material. As a rule of thumb we should keep in mind that high frequencies will be better reflected at material with low density and vice versa. Frequencies of 10–50 kHz are best reflected by rocks and very hard sediments. Transducer emitting frequencies of 100 kHz are used to observe the upper surface of sands and with 200 kHz they will receive the signal from soft sediments and also sea grass. Signals of higher frequencies of up to 1,000 kHz are reflected by particles suspended in the water and used for acoustic Doppler current profiling (ADCP). The fact of the frequency-depending reflectivity is used in the latest developments of hydroacoustic components. Newly developed sub-bottom profilers provide a complete tomography of the seafloor.

The resolution and the quality of depth measurements depend on the width of the beam angle that is transmitted by the transducer. A larger beam angle increases the possibility of depth errors due to reflections from the beam's edge. Therefore hydrographic echo sounders should have beams with angles less than 30°. Common transducers produce beam angles between 2 and 30°. The beamwidth depends on the diameter of the transducer and the emitted frequency. For a circular transducer, the beamwidth (in degrees) is roughly given as

$$\beta \approx 65^\circ \times \frac{c_w}{d} \times \frac{1}{f}$$

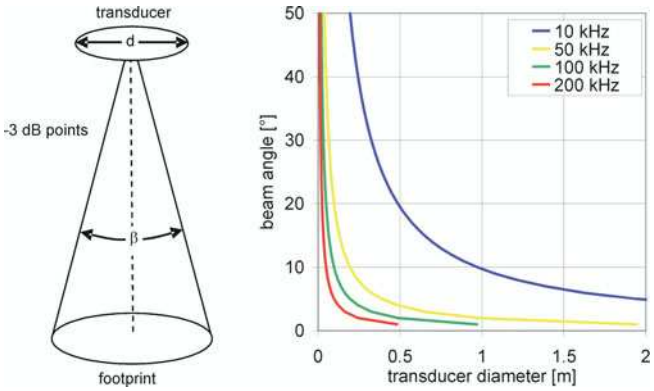


Fig. 8.6 Beamwidth vs. diameter

The higher the frequency the narrower the beamwidth for a given transducer size (Fig. 8.6). If a transducer is to transmit a sound with a frequency of 10 kHz and the beam angle is to be kept below 30°, the transducer's diameter must not be less than 35 cm. For higher frequencies the transducer's size can be kept much smaller.

8.2.2.2 Singlebeam Echo Sounders

Singlebeam echo sounders produce only one single beam that is commonly sent vertically into the water. They are frequently used to obtain depths directly under the vessel, thus avoiding wide-beam biases caused by underwater slopes. This depth is used either for safety of navigation or for seafloor mapping. Larger depths have to be corrected for roll and pitch motion of the vessel observed by a suitable kind of heave–roll–pitch sensor.

The beamwidth of conventional singlebeam echo sounders is usually of the order of 30°. However, narrow beam echo sounders with beam angles of less than 5° have also been available since the mid-1980 s. To produce a narrow beam, larger size transducers are needed than for a wide beam as can be seen from Fig. 8.6. The equipment becomes bulky and expensive (De Jong 2002).

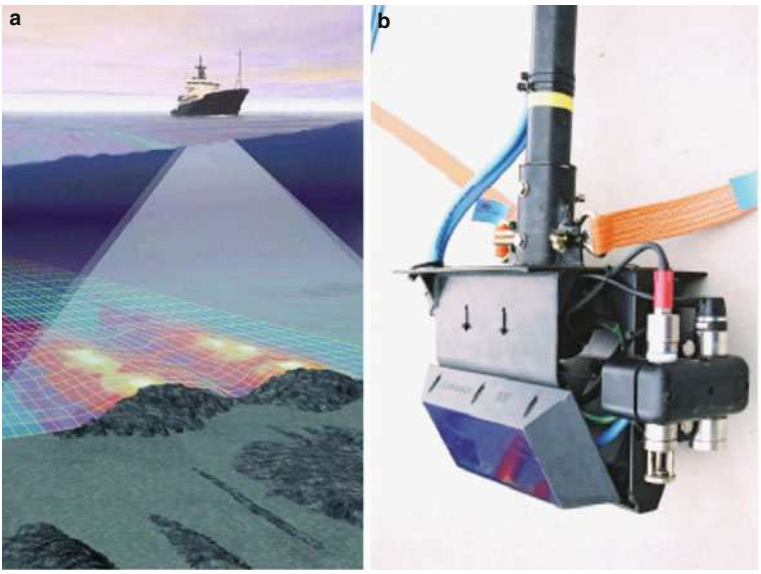
Singlebeam sounders (Fig. 8.7) are frequently used as two-frequency systems, e.g. 30 and 200 kHz. This provides good checking possibilities and additional data even in areas with large sediment transport and muddy seafloor like tide-influenced estuaries.

In comparison to multibeam sounders the purchasing costs of survey echo sounders with a single beam are much lower. They produce a smaller and easier to handle amount of data than multibeam sounders and are less influenced by roll and pitch motion in shallow waters. Especially for high-quality surveys, singlebeam echo sounders are still the most important and most widely used tool in hydrography.

Fig. 8.7 Singlebeam echo sounder (*left*), transducer (*middle*, Reson) and record and display unit (*right*, Fahrentholtz)

8.2.2.3 Multibeam Echo Sounders

Multibeam echo sounders increase the bottom coverage by using a fan of beams or swath instead of a single beam. Consequently they increase the productivity of hydrographic and bathymetric observation cruises. Particularly in large-scale bathymetric surveys, where accurate depth measurements for many neighbouring points on the seafloor have to be obtained so that an exact picture of the bottom profile can be established, this technique is much more appropriate than singlebeam sounding (Fig. 8.8a).



(DEOS, TU Delft).

Fig. 8.8 (**a**) Swath of a multibeam echo sounder and (**b**) transducer field of a multibeam echo sounder

For a reliable result the beams have to be formed and received in a precise way that allows the clear identification of the direction of the beam. Beams are commonly formed by a field of transducers which produce overlapping signals. In the overlapping area points of constructive and destructive interference are created. The signals are received by an array of hydrophones which only receive signals created at the constructive interference points. Due to the different phases of the received signal from different directions, the signals from different hydrophones add destructively. Collectively, the hydrophone array is not sensitive to sounds from this direction (Fig. 8.8b) (L-3 2000).

Each of the narrow beams produced yields a resolution of the bottom equivalent to that of a narrow singlebeam echo sounder. The measurement accuracy is not better than that of singlebeam echo sounders, however. In fact, accuracy decreases as the swath angle increases (De Jong 2002). The available frequencies range from 100 to 450 kHz (Blacquiere and van Woerde 1998). These frequencies lead to smaller transducer sizes that are necessary for a small field of transducers. The fan is narrow in the fore and aft direction, typically 1.5°, and wide in the port and starboard direction with a typical swath angle of 120° and up to 160°.

Due to the wide swath angle the beams at the outer end of the swath are very sensitive to attitude changes of the survey vessel. Therefore very good information on the present roll, pitch and yaw angle either from INS- or GPS-based attitude sensors are necessary. Simple and cheap attitude systems often fail during passages with slight changes of sailing direction or curved tracks, and consequently the attitude sensor should be carefully chosen. Although the outer beam data are often neglected in data processing, not all remaining errors resulting from uncertain attitude angles can be eliminated.

8.2.2.4 Side-Scan Sonar

A side-scan sonar reveals information about the seafloor composition by taking advantage of the different sound absorbing and reflecting characteristics of different materials and bottom shapes. Reporting the strength of echoes is essentially what a side-scan sonar is designed to do.

In contrast to a singlebeam echo sounder, a side-scan sonar records not only the reception time of the first echo but also the reception times of the following echoes. Additionally, the amplitude of the received signal is registered together with its reception time. If the amplitude is plotted against the time of reception (Fig. 8.9a), a shaded image of the ground and all objects in the water is produced (Fig. 8.9b).

A single signal sent into the water will propagate as a spherical pulse front. Thus, only the range of a reflection can be observed but not the direction to the reflector if only one single transducer is used. Most side-scan sonars deal with this problem by introducing some directivity into their projected pulses. This is done by using a line array of transducers. The long axis of the line array is oriented parallel to the direction of travel of the survey vessel (L-3 2000). The transducer array is commonly

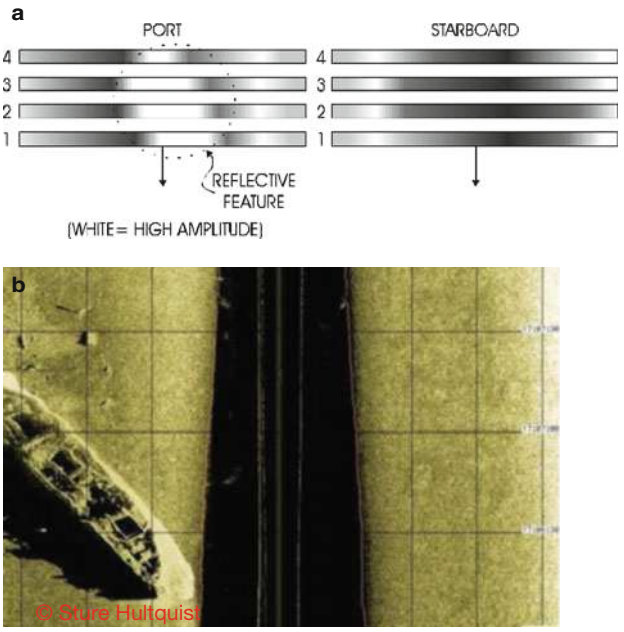


Fig. 8.9 (**a**) (*top*) Amplitude vs. time plot and (**b**) (*bottom*) *shaded image* of the ground and objects

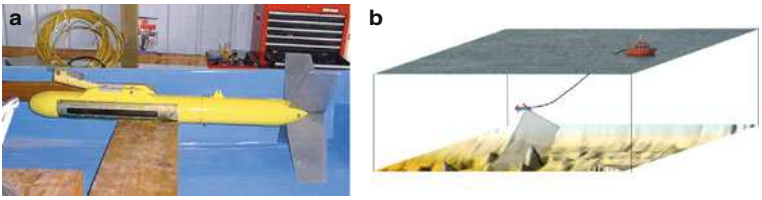


Fig. 8.10 (**a**) Towfish and (**b**) Towfish towed in the water

installed in a towfish (Fig. 8.10a) which is towed in the water somewhat below the surface – ideally in a height of 10% of the range above the bottom – behind a survey vessel (Fig. 8.10b).

8.2.2.5 Sub-bottom Profilers

Parametric echo sounders using non-linear acoustics are nowadays used to profile the sub-bottom of the sea. Basically a low-frequency signal will penetrate the soil on the seafloor deeper than a higher frequency signal. Unfortunately, very large transducer sizes are necessary to form a narrow beam on low frequencies.

Parametric echo sounders transmit two signals of slightly different high frequencies at high sound pressures (primary frequencies). Because of non-linearities in the sound propagation at high pressures, both signals interfere and new frequencies arise (Fig. 8.11). Due to the use of high-frequency signals the transducer size can be kept small also for a narrow beam (Wunderlich and Müller 2003).

The so-called secondary frequency (difference of the transmitted frequencies) is low and penetrates the sea bottom. The primary frequencies may be used for exact determination of water depth even in difficult situations. Modern systems are able to work in water depth up to 1,500 m with a penetration depth of 50 m. Commonly, primary frequencies of 100 kHz are used and low secondary frequencies between 4 and 15 kHz are produced with very small beam angle of up to 2°.

Sub-bottom profilers can be used in a manifold of applications. Figure 8.12 shows the result from a shallow water survey with the task to detect an embedded pipeline.

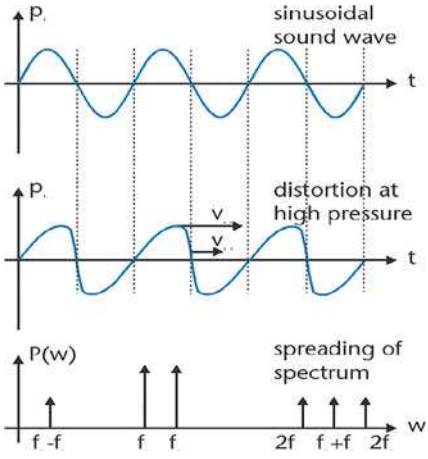


Fig. 8.11 Signal interference of a sub-bottom profiler (Wunderlich/Müller)

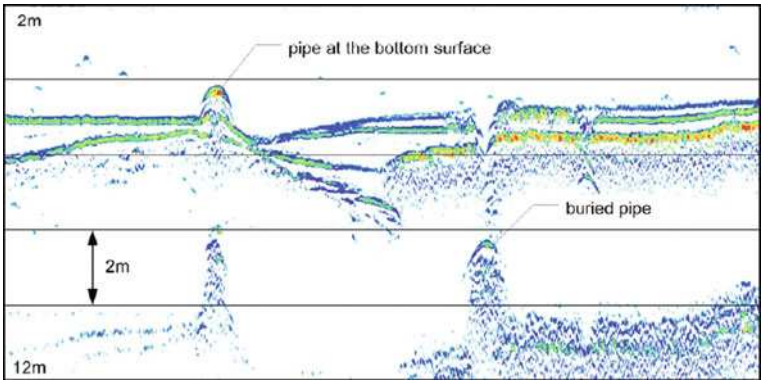


Fig. 8.12 Shallow water survey using sub-bottom profiler (Wunderlich/Müller)

The pipe diameter was 30 cm and the pipe was detected over the whole area under a covering layer of 1–2.5 m.

8.3 Precise Navigation

The contribution of marine geodesy to precise navigation is to ensure safe and efficient navigation of ships in open waters as well as in coastal areas and approach channels. The manifold applications cover the scopes of simple positioning by means of GPS, the production and use of analogue and digital maps, the navigational-oriented GIS application, ships attitude determination and investigations on ship's hydrodynamics.

8.3.1 Maps of Coastal Waters and Approach Channels

Although the development of electronic navigation systems has reached a very high level, classical paper charts are still in use on every ship as the basis of navigation. The International Maritime Organization (IMO) regulation requires all ships to carry nautical charts to plan and display the ship's route for the intended voyage and to plot and monitor positions throughout the voyage. The ship must carry paper charts at least as a back-up arrangement if electronic charts are used.

A nautical chart is a graphic representation of the marine environment showing the form of the coast, the general configuration of the sea bottom including water depths, locations of hazards to navigation, locations and characteristics of man-made aids to navigation and other features useful to the mariner. Additionally, special "roads", such as traffic separation schemes, in areas with high volume of traffic are exactly displayed (Fig. 8.13).

Nautical charts are produced by national hydrographic services which are responsible for all charts covering the EEZ of a state. Many national services also publish maps of other national and international waters so that some areas are covered by charts from different hydrographic services. Nowadays some hydrographic services like NOAA offer nautical charts as print-on-demand (POD) version. They are equivalent to the traditional paper charts with the significant enhancement that the charts are updated continuously with Notice-to-Mariner corrections.

8.3.2 ENC and ECDIS

Electronic navigational charts (ENC) are vector-based digital files containing marine features suitable for marine navigation. They are commonly based on the International Hydrographic Organization (IHO) S-57 standard. Similar to paper charts, ENCs are produced by national hydrographic services (Fig. 8.14a).

ENC are used in Electronic Chart Display and Information System (ECDIS). Technically similar to GIS – used for many applications on land – ECDIS is a

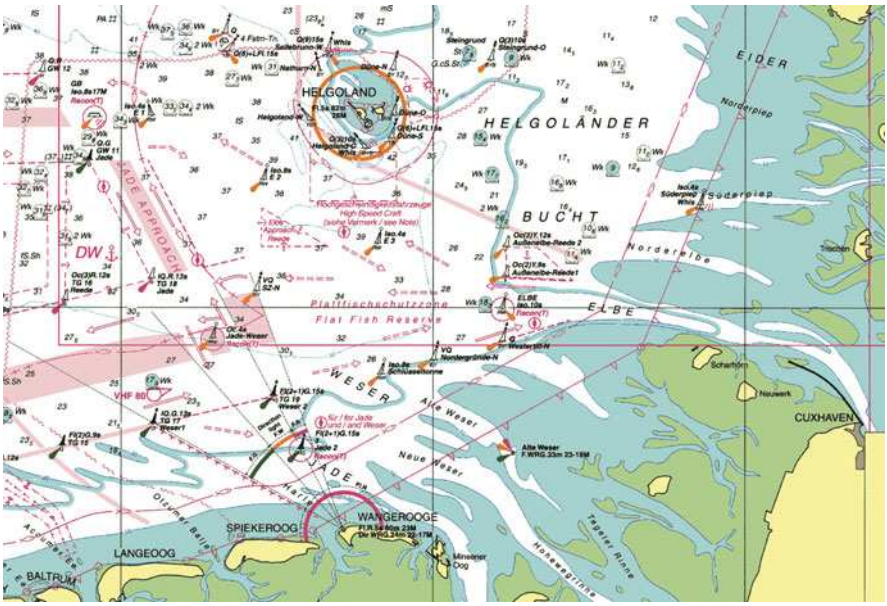


Fig. 8.13 Example of nautical map (BSH Chart 87, Int 1413, 2004)

computer-based navigation information system that complies with International Maritime Organization (IMO) regulations. The system displays ENCs and integrates information from GPS and other sources such as radar and automatic identification systems (AIS). It may also display additional navigation-related information, such as waypoints, distances and bearings (Fig. 8.14b).

The primary function of an ECDIS is to contribute to safe navigation. It combines different safety-relevant information from a variety of sources in a single display system and, if applied correctly, can warn the user of dangerous situations.

8.3.3 Ship's Attitude

Precisely determined heading, roll and pitch angles are primary parameters for several scientific instruments and can also be essential to safe navigation. Ship motion measurement systems are used in many applications. Helideck monitoring systems, for example, are developed to provide full information about the helideck movements to both heliports and land-based operators.

Other systems are used for compensation of vessel motions to satellite dish stabilization in rough weather conditions. Ship's attitude information is the basis of automatic control systems of marine vehicles particularly used on supply ships for oil platforms ("dynamic positioning"). Up-to-date systems for real-time wave

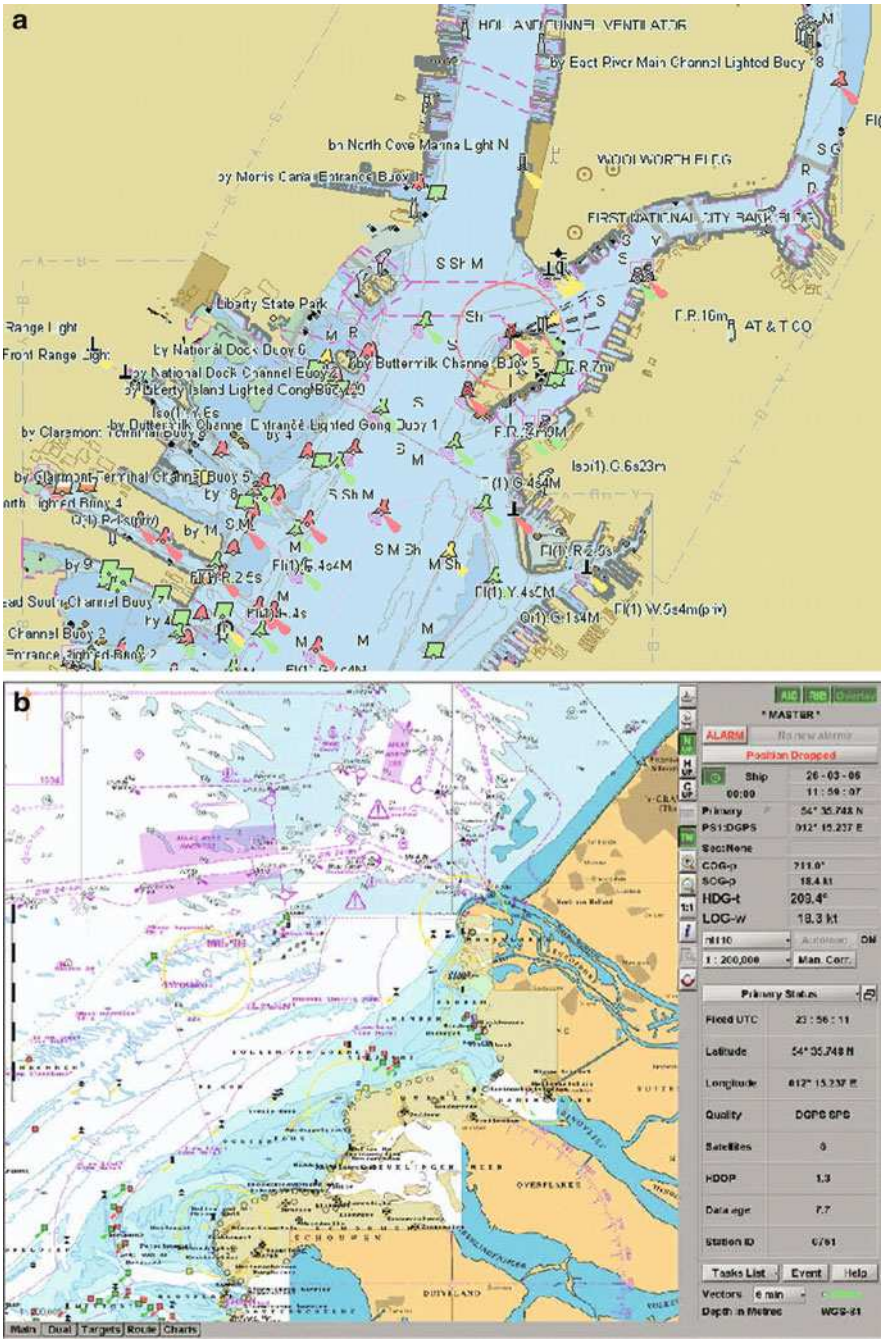


Fig. 8.14 (**a**) (*top*) ENC of ports of New York and (**b**) (*bottom*) ECDIS display of northern Norway (reproduced from NOAA'S National Ocean Service Chart 12345, not for use in navigation, Fugawi View ENC © 2007 Northport Systems Inc.)

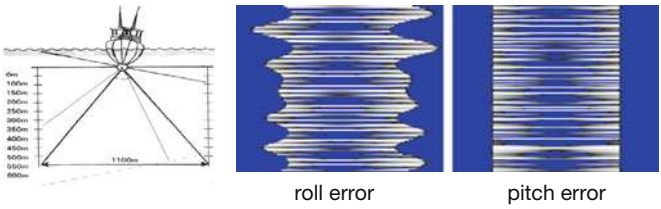


Fig. 8.15 Effects of roll and pitch errors on multibeam echo soundings

monitoring (Nieto Borge et al. 2004) can also be used on moving platforms if precise attitude information is available. Those systems might be able to support ship routing systems that are used to define best and safest passages.

In bathymetry and hydrography, multibeam echo sounders require real-time attitude data as an input to rotate the frame of reference of their measurements from ship-relative to earth coordinates (Fig. 8.15). Roll and pitch frequencies of small vessels are of the order of 1/10 Hz or higher. The measurement rate of a swath system depends on the water's depth and is commonly of the order of 20–40 profiles per second. Consequently, the attitude determination system should be able to provide the roll and pitch information with a frequency of at least 10 Hz or better.

In the past, INS were generally used to derive the ship's attitude because they were able to produce data sets with a frequency of typically 100 Hz. Since the advent of precise GPS-on-the-fly solutions based on real-time phase observation with sample rates of up to 0.05 s, more and more GPS-based attitude systems have been used to derive corrections for multibeam observations (Andree et al. 2000). The advantage of GPS-based systems is that their solutions are less deteriorated by disturbing accelerations, e.g. from following curved tracks.

8.3.4 Hydrodynamics of Ships

The contribution of marine geodesy to scientific investigations in ship's hydrodynamics is based on full-scale observation of the ship's behaviour under particular hydrodynamic conditions by means of precise GPS and additional hydraulic parameters, i.e. waves, tides, currents and seafloor morphology and topography.

During the last decade, significant progress has been achieved, especially in full-scale squat observation and dynamic trim measurements. The results of these observations are used as an important input to the description of a ship's behaviour in confined and open waters. Recent developments show that data derived from marine geodesy will be used in real time under-keel-clearance estimation of ships and might also improve the efficiency of ships, waterways and harbours (O'Brien and O'Brien 2004).

8.3.4.1 Basics of Squat

According to the principle of Archimedes, an object partially submerged in a fluid displaces a volume of fluid that weighs the same as the apparent loss in weight of the object. This means that the draught of a ship in static equilibrium is directly related to its displacement (Fig. 8.16a). Depending on the distribution of ballast water and cargo, the draught of a ship might be greater at bow (negative trim) or stern (positive trim) or equal at the forward and aft perpendiculars (trimmed on even keel).

As a ship starts to move through water she creates a system of streamlines around her hull which in turn, according to Bernoulli's equation, change the distribution of pressure. This simple picture is enough to understand how, in principle, the bow and stern waves and the trough alongside a moving ship are generated. The immersion of the ship will still be according to its static draught and trim but now the ship is affected by its own wave system. Due to the lowering of the surrounding water level in comparison to the undisturbed surface there is an apparent increase of draught and trim. The change of draught and trim of the moving ship with respect to the unperturbed water level is called ship squat (Fig. 8.16b).

As the cross section available for the streamlines is reduced in shallow waters and narrow channels, the water alongside the ship is accelerated more intensely and the squat effect is more pronounced. Awareness of squat in the shipping community has been roused in recent decades as more and more large size ships use restricted waterways to their limits. It is not uncommon in these cases to allow for a squat effect of more than 1 m.

Since the squat reaches values of this scale it is a serious influencing factor particularly for traffic of merchant ships. From the point of view of security it has to be taken into account very carefully to avoid groundings. Port and waterway authorities commonly endeavour to guarantee the highest possible safety level for passages by restricting transits, especially in waterways with tidal influence. Additional draught restrictions for very large ships are also common. In an economic perspective the draught of a ship should be as large as possible for an efficient use of the capacity of waterways and harbours. To ensure a safe and efficient use of the entire maritime infrastructure the squat of a ship has to be known and predicted up to a scale of some centimetres.

Essentially, the amount of squat depends on speed-through-water, size and shape of the ship's hull and the cross section of the waterway. Detailed numerical calculations based on methods of computational fluid dynamics (CFD) are numerically

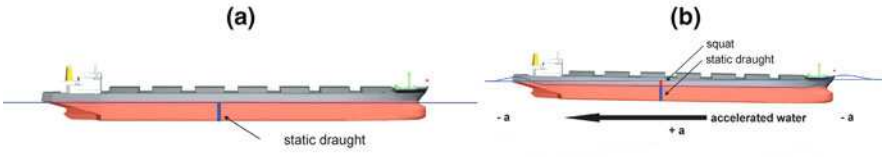


Fig. 8.16 (**a**) Ship in static state and (**b**) moving ship and accelerated water

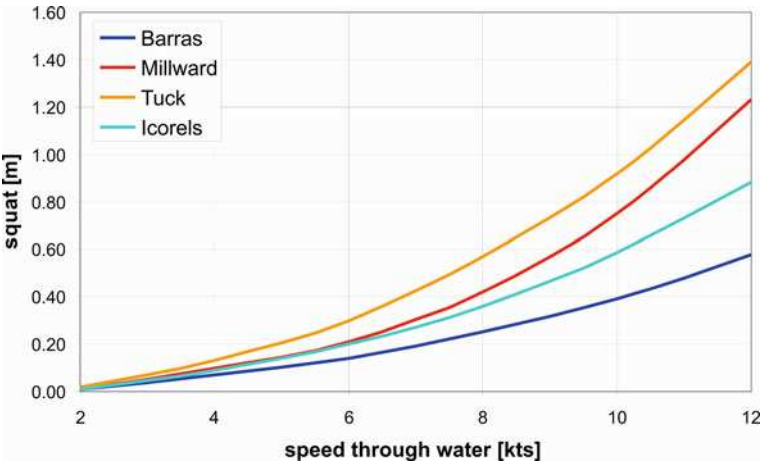


Fig. 8.17 Squat derived from different empirical formulae

very demanding and yield a result only for one particular stationary condition. A rather successful analytical approach has been made by Tuck (1966). Yet, for practical purposes, it is customary to use empirical formulae (PIANC-IPAH 1997) that were developed in the 1970 s from few experimental data available then. Some of these formulae are only valid for restricted ranges of parameters and, in cases where comparisons are possible, results differ substantially (Fig. 8.17).

It is easy to see that empirical formulae have to be improved for a reliable use on board a ship. Enhanced formulae can only be expected if high-quality data sets from full-scale experiments are used as the basis for investigations.

8.3.4.2 SHIPS Method

Since the advent of high-precision satellite navigation the direct measurement of squat has become more promising. In 1998 the SHIPS (SHore-Independent Precise Squat observation) method was introduced by the University of Applied Sciences in Oldenburg (Härting and Reinking 1999) and has been improved since (Dunker et al. 2002). The principle of the SHIPS method is illustrated in Fig. 8.18.

Three GPS receivers are operated on the ship such that, apart from the overall vertical movement which will be described by the height change of a ship's reference point (LCF, longitudinal centre of floatation), changes in trim and list can also be observed. Another GPS receiver is installed on a small escort craft travelling ahead, outside the ship's wave system. The purpose of the escort craft is to represent the unperturbed water level at the measurement position. The receiver on the escort craft is used as a mobile reference station and the carrier-phase DGPS solution for the receivers on the ship is computed directly with respect to the escort craft. Thus, with the SHIPS method, the squat is determined independent of tide gauges and

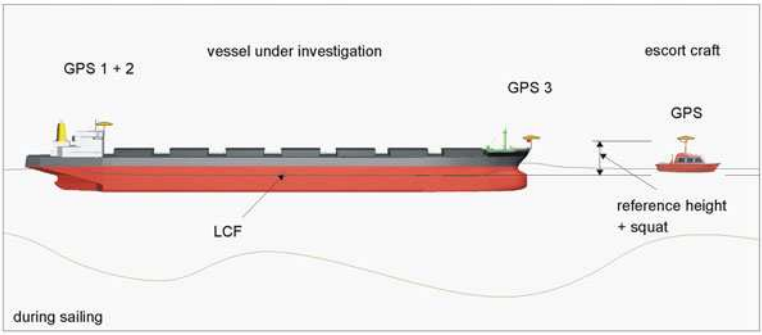


Fig. 8.18 Principle of SHIPS method

shore-based reference stations, assuring a high-quality carrier-phase DGPS solution by a short baseline between receivers, even if the experiment is run over large distances.

To derive squat information from observed height changes and to allow a proper data analysis, some corrections must be taken into account:

1. The speed-over-ground is obtained from GPS observations with adequate accuracy. The speed-through-water and current information is obtained from ADCP observations on board the escort craft.
2. The draught of a ship depends on the density of the water. Therefore samples of salinity and temperature have to be taken during the experiment, especially if they are carried out in estuaries.
3. In areas with tidal influence the water surface gradient must be derived from an average tide wave model or from tide gauge readings.

Additionally, the height of the escort craft is influenced by the squat behaviour of the boat and by its wave- and swell-induced heave. To determine the speed dependence of the GPS antenna height on board the escort craft, a calibration experiment has to be carried out. The boat can be driven at various speeds, stopping (and turning) in between in the vicinity of a land-based reference station or better a reference station on a floating platform while GPS observations are performed. Figure 8.19a shows an example data set of a calibration experiment with variable speeds (lower graph) and resulting height changes (upper graph) plotted over seconds of GPS day. Since the manoeuvres of the small boat require only a few seconds, the height differences between stationary speed and drifting can be assigned to the "squat" at that speed (Fig. 8.19b).

The wave-induced height variations in hydrographical measurements are frequently corrected for by using heave–roll–pitch sensors. However, if such a sensor is not installed on the escort craft, height variations can also be determined using precise onboard GPS receivers. For this, height changes between successive epochs are computed by epoch-to-epoch GPS–phase differences (Reinking and Härting

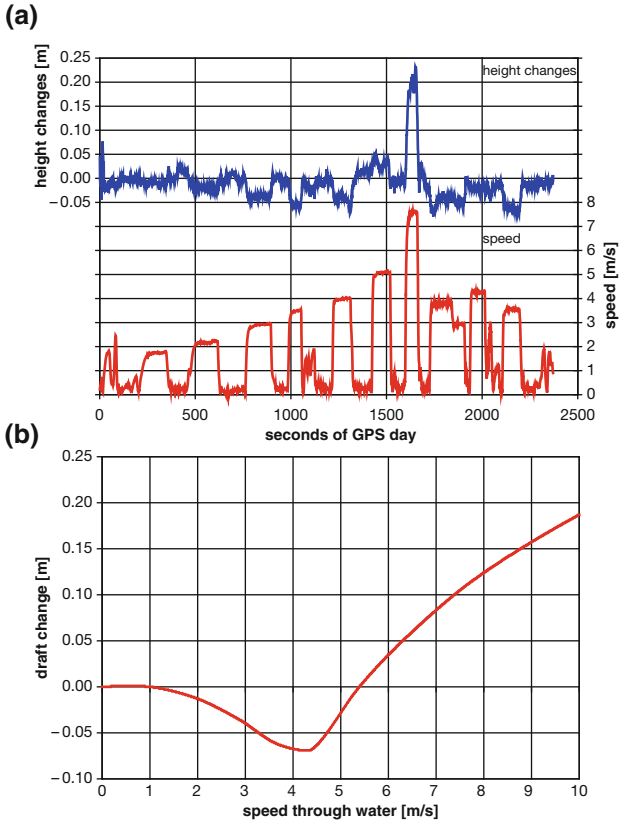


Fig. 8.19 (**a**) (*top*) Data from calibration and **(b)** (*bottom*) resulting calibration function

2002/1). This is easily done under the assumption that atmospheric effects can be neglected if the time difference between the epochs is short. The vertical movement can be calculated by cumulating the height changes but unfortunately the neglected atmospheric effects are leading to artificial long-term variations. Tests have shown that the application of a sixth-order high-pass Butterworth filter with a 20-s cutoff period reduces or eliminates these long-term variations without spoiling the information about the vertical movement induced by short-period waves. The result is a high-quality wave correction that fits exactly to GPS observation epochs and positions (Reinking and Härting 2002/2) (Fig. 8.20).

In total, 18 observation campaigns have been conducted during the last few years using the SHIPS method. The results have clearly demonstrated that the ship's squat can be derived with an accuracy of about 2–4 cm by this method. An experiment of particular interest was carried out in October 2004 at the Rio de La Plata starting in Buenos Aires on the outgoing journey of the container vessel *CAP FINISTERRE*. The Rio de La Plata estuary shows an average depth of 4–5 m but the dredged approach channel to Buenos Aires has a maintained depth of 10.1 m. Only very few

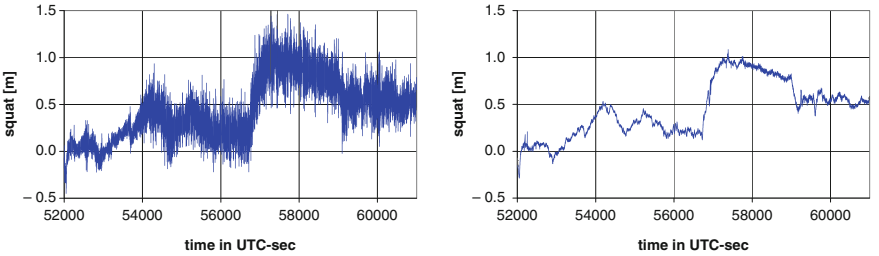


Fig. 8.20 Squat results without (*left*) and with (*right*) application of GPS-derived heave correction (La Plata experiment: wave height approx. 1 m)

tide gauges and land-based GPS reference stations are available in this area so that in this region the SHIPS method is unbeatable.

The experiment was driven up to pilot change at Ensenada on a total trip distance of about 48 km. Due to technical problems of the escort craft which was contributed by the Servicio Hidrografia Naval of Argentina, the trial had to stop before reaching the expected end of the test track at Punta Indio, 130 km away from Buenos Aires. The outcomes are plotted over GPS time in seconds and presented in Fig. 8.21 with comparable results from squat calculations using different empirical formulae.

It can clearly be realized that none of the empirical formulae is able to describe the squat behaviour correctly; in particular the performance at 59,000 s is completely ignored by these squat functions. The reduction of speed at this stage

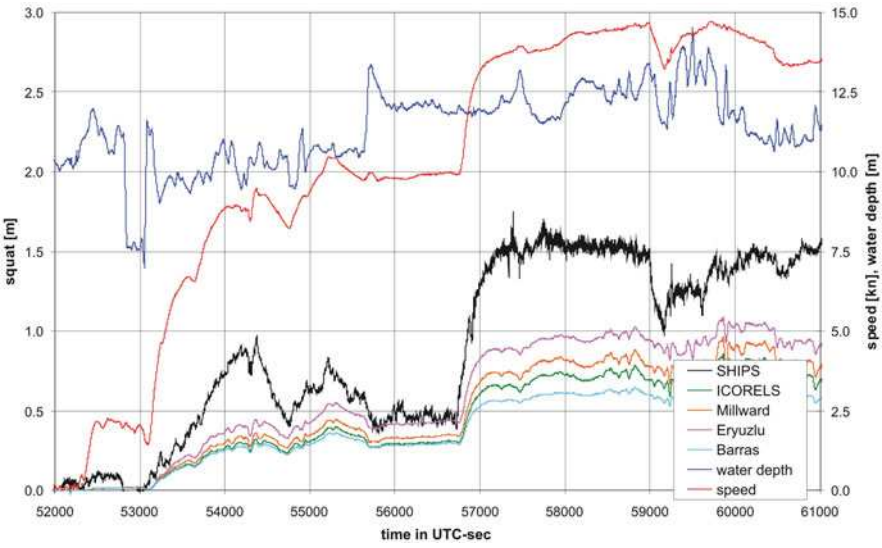


Fig. 8.21 Results from squat experiment on Rio de La Plata

could be a result of the reduced under-keel-clearance leading to a squat reduction and not influenced by manoeuvres on board the vessel under investigation. Dynamic processes similar to this are not predictable by means of simple squat formulae.

8.3.4.3 Squat and Trim

Taking the results of the above-mentioned experiment and plotting the squat over the dynamic trim angle, which is not only a result of hydrodynamic forces but also influenced by pitch and acceleration effects, a clear correlation between squat and trim can be found (Fig. 8.22).

The present dynamic trim of a ship can be observed in real time by means of low-cost GPS-phase observations up to an accuracy level of some centimetres. Consequently, the question arises as to whether it would be possible to observe and use the dynamic trim as an additional parameter for the improvement of empirical squat formulae. This question is not easy to answer and needs some additional investigations, particularly on the interaction between the ship and the detailed morphology of the waterway.

Future research will have to use 4D modelling of the observations from full-scale experiments including the detailed seabed structure and the ship's actual motion to compare these findings in parts of particular interest with results from complex CFD calculations.

It is a well-known fact that the squat also depends on the static trim of a ship. Yet it is not completely understood whether a suitably selected static trim leads to a reduction of squat or not. Some results from experiments in Germany show that a positive static trim reduces an additional bow squat for specific ship in particular waterways (Fig. 8.23).

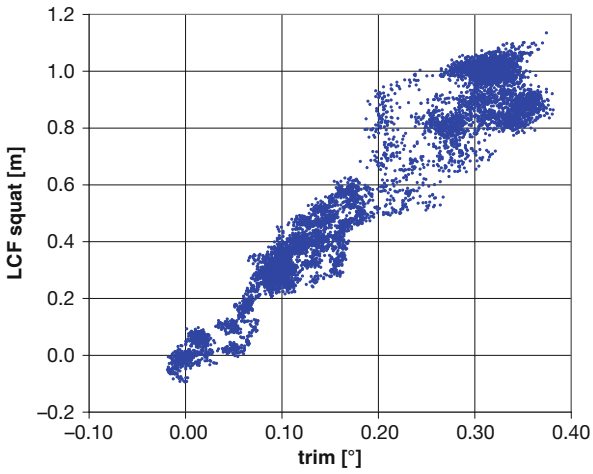


Fig. 8.22 Trim and squat of La Plata experiment

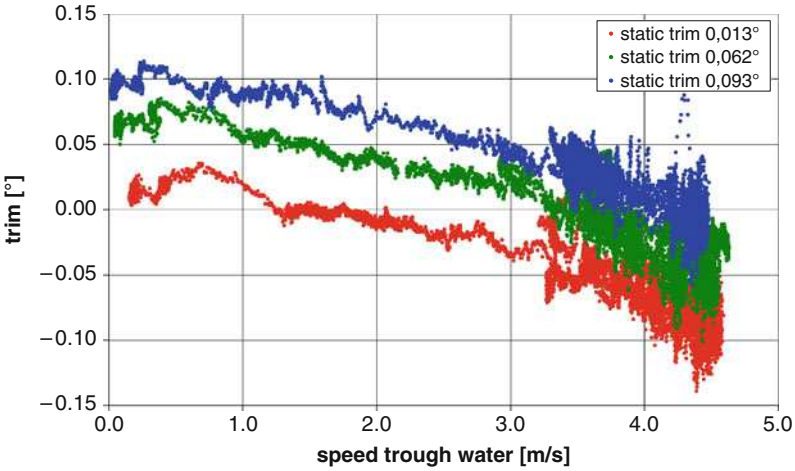


Fig. 8.23 Trim over speed through water from three different experiments

Upcoming researches will focus on the question of how the squat can be optimized by a well-chosen static trim. Additional full-scale experiments are foreseen in which the static trim will be steered in a safe but significant manner. It is envisaged to generalize the resulting facts for a group of ships and to set up a set of rules for static trim recommendations that might be used as a standard on board a ship.

8.4 Conclusion

Marine geodesy has a very wide scope of work with a large variety of activities. It has become a nearly independent field of research with relations to many other areas of geodesy and influences from geology, geophysics, glaciology and navigation. The results of marine geodesy contribute to numerous fields like politics, economics and ecology.

We should not forget that all the work done by people involved in marine geodesy is not only interesting from a scientific point of view but marine geodesy is also a good business area for geodesists.

References

Andree, P., Läger, R., Schmitz, M. and Wübbena, G. (2000) Bestimmung von Schiffsbewegungen und anderen hochfrequenten Bewegungen mittels GPS, DGON Symposium Ortung + Navigation 2000, www.geopp.de/download/dgon2000.pdf, last access October 2006

Blacquiere, G. and van Woerde, K. (1998) Multibeam Echosounding: Beamforming Versus Interferometry. In: Oceanology International 98: The Global Ocean. Spearhead Exhibitions

Ltd., Brighton, www.elac-nautik.de/web/site/pdf/vertech/beamforming.pdf, last access October 2006

De Jong, C.D., Lachapelle, G., Skone, S. and Elema, I.A. (2002) Hydrography. Delft University Press, Delft, ISBN: 9040723591

Dunker, S., Gollenstede, A., Härting, A. and Reinking, J. (2002) Analysis and comparison of SHIPS-derived squat. Proceedings Hydro 2002, Kiel, pp. 434–449

Härting, A. and Reinking, J. (1999) Efficient determination of ship squat: first results. Schiff & Hafen, 3, 78–81

International Hydrographic Organization (IHO) (1998) IHO Standards for Hydrographic Surveys, 4th Ed. Special Publication No. 44, April 1998

L-3 Communications SeaBeam Instruments (2000) Multibeam Sonar Theory of Operation, www.ldeo.columbia.edu/res/pi/MB-System/formatdoc/L3TheoryOperation/thop_toc.pdf, last access October 2006

Ladage, S., Weinrebe, W., Gaedicke, C., Barckhausen, U., Flüh, E.R., Heyde, I., Krabbenhoeft, A., Kopp, H., Fajar, S. and Djajadihardja, Y. (2006) Bathymetric survey images structure off Sumatra. Eos Trans. Am. Geophys. Union, 87(17), 165, 168

Nieto Borge, J.C., Rodriguez Rodriguez, G., Hessner, K. and Izquierdo Gonzalez, P. (2004) Inversion of marine radar images for surface wave analysis. J. Atmos. Ocean. Technol., 21, 1291–1300

O'Brien, T. and O'Brien, P. (2004) UKC management through dynamic underkeel clearance systems. In: Härting, A. and Reinking, J. (eds) 2. Squat-Workshop 2004: Aspects of Underkeel Clearance in Analysis and Application, Elsflether Schriften zur Seeverkehrs- und Hafenwirtschaft, ISSN 0949-3530, Elsfleth 2004

PIANC-IPAH (1997) Approach channels, a guide for design. Report of the Joint Working Group II-30, PIANC-Bulletin No. 95

Reinking, J. and Härting A. (2002/1) Heave determination by stand-alone GPS and/or inertial sensors. Proceedings Hydro 2002, Kiel, pp. 452–459

Reinking, J. and Härting, A. (2002/2) GPS-gestützte Seegangskorrektur hydrographischer Messungen aus Einzelempfänger-Daten, Z. f. Verm.wesen 127, Heft 3

Schenke, H.-W., Dijkstra, S., Niederjasper, F., Hinze, H., Hoppmann, B. and Schöne, T. (1998) The new bathymetric charts of the Weddell sea: AWI BCWS. In: Jacobs, S.S. and Weiss, R.F. (eds) Ocean, Ice, and Atmosphere: Interactions at the Antarctic Continental Margin. American Geophysical Union (Antarctic Research Series, 75), Washington, pp. 371–380

Tuck, E.O. (1966) Shallow-water flows past slender bodies. J. Fluid Mech., 26(part 1), 81–95

UNCLOS (2006) United Nations Convention on the Law of the Sea of 10 December 1982, Overview and full text, www.un.org/Depts/los/convention_agreements/convention_overview_convention.htm, last access October 2006

Wunderlich, J. and Müller, S. (2003) High-resolution sub-bottom profiling using parametric acoustics. Int. Ocean Syst., 7(4), 6–11, July/August 2003

Chapter 9
Satellite Laser Ranging

Ludwig Combrinck

Contents

L. Combrinck (✉)
Space Geodesy Programme, Hartebeesthoek Radio Astronomy Observatory, PO Box 443, Krugersdorp 1740, South Africa
e-mail: ludwig@hartrao.ac.za

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_9,

9.1 Background

Determining the range to a satellite in orbit around the earth utilising the technique of satellite laser ranging (SLR) was pioneered in the early 1960s. The first successful ranging experiment was reported in the 3 December 1964 issue of Flight International (Smith 1964). Dr. Henry H. Plotkin of Goddard Space Flight Centre led a NASA team to track the Beacon-B (also known as Explorer-22) satellite for ten successful sessions during the period 11 October to 13 November 1964. A team from General Electric Co. (Valley Forge, Pennsylvania) also participated from Phoenix, Arizona. Using a telescope mounted with a ruby laser, expected range accuracy was about 3 m. Current accuracy is at the level of 1–2 cm.

9.1.1 Introduction

Future applications of SLR were recognised quickly after the initial success, and an obvious supplementation and future replacement application of SLR involved the network of Baker–Nunn cameras (Henize 1957), which had an effective limiting magnitude of about 16 for 30-s exposures (Solomon 1967). These cameras were used to track satellites and record rocket stages firing amongst other applications.

Results from a series of experiments throughout 1966, at the Smithsonian Astrophysical Observing Station, Organ Pass, New Mexico, using a pulsed ruby laser, indicated how lasers may be used to supplement the then worldwide network of Baker–Nunn cameras in obtaining precise satellite orbits for geodesy and other purposes (Anderson et al. 1966). The network of cameras was subsequently phased out during the late 1970s as SLR stations were established during the 1970s. Some of these decommissioned Baker–Nunn cameras were utilised for other suitable functions, such as comet hunting and photometry.

The 1960s radar observations of satellites as well as observations using the Baker–Nunn network allowed station positions to be determined, but these positions were only accurate to within about 100 m. In order to exploit satellite orbital information to determine station position accurately enough, so as to enable the evaluation of models which calculated dynamic processes of the Earth which affected station positions such as Earth-tides and plate tectonic motion, the accuracy of SLR measurements would have to be at the centimetre level. These dynamic processes displace station positions at the centimetre level from sub-diurnal to continuous time scales. The success with ranging to the Beacon satellite led to the launch in 1965 of GEOS 1 (Geodetic Earth Orbiting Satellite) by NASA. This satellite was exclusively designed for geodetic studies and was equipped with SLR reflectors as well as a radio range transponder, a range and range rate transponder, Doppler beacons and four optical beacons. GEOS 1 was tracked by Smithsonian Astrophysical Observatory and NASA SLR stations. The array of instrumentation on the satellite was used to determine three-dimensional station positions with an accuracy of 10 m, Earth's gravity field structure and also improvements of the locations and magnitude of gravity anomalies. In addition, an inter-comparison of the accuracy of the various systems equipped on the satellite was done.

Fig. 9.1 LAGEOS, a 60-cm diameter sphere covered with 426 retroreflectors. Launched on 4 May 1976, it will remain in its near circular orbit for many decades. Source: NASA

Not only is SLR a valuable tool to measure these small centimetre-level station position variations resulting from geophysical processes, but from early on it provided important contributions to the development of gravity models of the Earth. For instance, SLR data from GEOS-1 (20 January 1977 through to 14 December 1978) were included in gravity model EGM96. Satellite laser ranging is a unique technique for observing the slow-varying geodynamic processes of Earth and the long wavelength components of the gravity field and their variation in time; satellites such as LAser GEOdynamics Satellite (LAGEOS, see Fig. 9.1) (Smith and Dunn 1980) have a very long lifetime and will be studied (and augmented with others) for many years to come, making it an ideal and stable long-term scientific tool. The SLR technique is a critical component in establishing an accurate global reference frame (Tapley et al. 1993). In addition, SLR contributes to an absolute scale factor for orbit determination, through using LAGEOS ranging data to determine the gravitational constant (Dunn et al. 1999). As SLR contributes (in conjunction with other techniques) to orbital calibration of satellites equipped with radar altimeters, it has and is making a valuable contribution to space oceanography (Bonnefond et al. 1995; Luthcke et al. 2003). New applications are being developed and the regular ranging horizon is being extended from Earth orbiting satellites to the Moon (one-way tests have been done to Mars Orbiter Laser Altimeter in 2005), with the launch of the Lunar Reconnaissance Orbiter (LRO) during 2009, which will be tracked (one-way ranging) with SLR for orbit calibration purposes (Smith et al. 2006); MOBLAS-6 at HartRAO will also participate in this historical event.

9.1.2 Basic Principles

The basic principles of SLR are quite simple. Figure 9.2 describes the main components of an SLR system. A satellite equipped with a corner cube reflector (or an

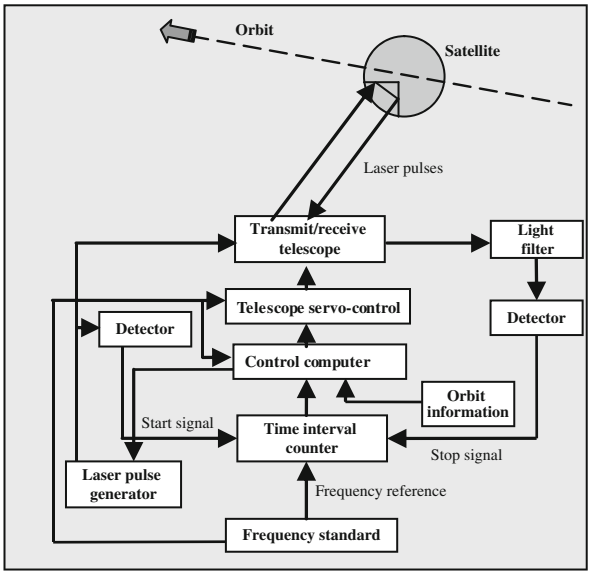


Fig. 9.2 Basic components of a satellite laser ranging system, describing signal paths

array of CCRs) is tracked by an optical telescope which has a sensitive light detector at its receiving end. In parallel and co-aligned, a transmit telescope emits short laser pulses at a rate of, say, 5 Hz. The departing laser pulses trigger an interval counter at a certain time (epoch of data); the laser pulses are reflected and the received laser light pulses are registered by a sensitive light-detecting device. The detector permits a stop signal to be sent to the interval counter. Utilising these round-trip time intervals (time-of-flight) and the speed of light, the range is half of this two-way range. Most SLR systems do require some form of human intervention to operate successfully. This is basically due to the fact that the combination of mechanical misalignment, clock offsets, instrumental and orbital biases as well as instrument-specific requirements creates a situation where the laser might be pointing slightly off target. The operator then has to make allowance for three offsets: along-track and cross-track biases, as well as time bias. These offsets are recorded on the computer and could serve as a starting point for subsequent satellite passes.

Apart from having to point in exactly the correct direction, other adequate conditions concerning laser power level, receiving telescope aperture and atmospheric conditions amongst others are required to receive successfully photons back from the satellite CCRs. The success of receiving returns can be estimated by the radar range equation (Degnan 1993), where the mean number of photoelectrons recorded by the SLR detector N_{pe} is given by

$$N_{\mathrm{pe}} = \eta_{\mathrm{q}} \left(E_T \frac{\lambda}{hc} \right) \eta_t G_t \sigma_{\mathrm{sat}} \left(\frac{1}{4\pi R^2} \right)^2 A_R \eta_r T_a^2 T_c^2. \tag{1}$$

In (1), η_q is the detector quantum efficiency (the fraction of the total radiation incident that is actually detected), E_T is the pulse power (average power divided by pulse repetition rate), λ is the wavelength of the laser, h is Planck's constant, c is the speed of light in a vacuum, η_t is the efficiency of the transmitter optics, G_t is the transmitter gain, the satellite optical cross-section is given by σ_{sat} and R is the slant range to the satellite. The effective area of the SLR telescope receiving aperture is A_R , η_r is the efficiency of the receive optics, T_a is the one-way atmospheric transmission and if cirrus cloud (thin, wispy cloud, occurring at altitudes >6 km, composed of ice crystals) is present, T_c is the one-way transmissivity of cirrus cloud. Slant range can be calculated (Degnan 1993) using the equation

$$R = -\left(R_E + h_l\right)\cos\theta_{\text{zen}} + \sqrt{\left(R_E + h_{\text{l}}\right)^2\cos^2\theta_{\text{zen}} + 2R_E\left(h_s - h_l\right)h_s^2 - h_l^2}. \quad (2)$$

Here the radius of the Earth is given by R_E and h_l and h_s are the altitudes of the station and satellite above sea level, respectively. The zenith angle of the satellite (complement of the elevation angle) θ_{zen} is as observed from the SLR station. A general expression for transmitter gain (Degnan 1993), which takes into account the effects of radial truncation of the Gaussian beam caused by some limiting aperture (such as the main transmitter primary) and central obscuration (normally caused by the secondary mirror of a Cassegrain telescope) is given by

$$G_t = \frac{4\pi A_t}{\lambda^2} g_t\left(\alpha_t, \beta, \gamma_t, X\right). \quad (3)$$

In (3) the transmitting aperture is given by $A_t = \pi a_t^2$ and $g_t\left(\alpha_t, \beta, \gamma_t, X\right)$ depends on geometrical factors such as whether the collimating telescope is perfectly focussed and whether the target is in the far field of the transmitter. More details are to be found in Klein and Degnan (1974) and Degnan (1993). Considering MOBLAS-6 and utilising these equations for a clear day with no cirrus clouds present and making some assumptions concerning good and bad weather conditions (atmospheric transmittance of 0.8 and 0.02, respectively), one finds that the maximum number of received photoelectrons could vary between 641 and 0.04 per pulse. This clearly illustrates the weather dependency of this technique. The location of an SLR station is therefore a critical factor; however, this matter seems to have carried less weight than other factors in the installation of several stations, e.g. collocation advantages with other space geodetic techniques as was the case with installing MOBLAS-6 at Hartebeesthoek Radio Astronomy Observatory in South Africa. If there are no other compelling reasons, a site with minimum cloud and high atmospheric transparency will yield better results and more data than an inferior site and should be given preference.

In Fig. 9.2, a very important subsystem is the frequency standard; this unit as well as the time and frequency distribution throughout the other subsystems of the SLR is extremely important. An example will make this clear; the velocity of a satellite frequently used for different applications of SLR, LAGEOS, during a test analysis of a 1-day arc (~6.3 orbital revolutions) indicated a minimum velocity of 5,645 m/s

and a maximum velocity of 5,806 m/s. This means that a timing precision of 1.7×10^{-7} or 0.17 μs is required to register the epoch of the observation. If the accuracy, i.e. how close the timing value is to the real value, is not exact, a small time bias can develop as the epoch of the observation can then be either too late or too early. Generally, SLR station time bias values are at the few microseconds level, and this is normally detected during post-analysis of the tracking data and corrected during the analysis. With regard to determining the range to the satellite, the velocity of light $\left(\sim 3 \times 10^8 \text{ m/s}\right)$ gives one a two-way range precision of 0.15 mm/ps (ps = picosecond), so that in order to reach a few millimetres, a precision of at least 20 ps should be reached.

Currently, the objective of the SLR community is to reach millimetre accuracy in ranging, indicating that an improvement factor of 10 will be required. Therefore new systems, e.g. the 1-m SLR/LLR system being developed by HartRAO (South Africa) in collaboration with OCA (France), will require timing systems (event timer or interval counter) with 1- to 2-ps precision. This is a very demanding requirement; even so, several high-precision timing systems are being developed which approach these accuracies (Artyukh 2007).

9.2 Range Model

Armed with the short background material of the previous section, this section will describe how the real range to the satellite can be determined. Table 9.1 gives a good overall view of some of the accuracies and precisions to be found in a modern SLR station. An average of 6- to 9-mm precision is achieved to a single CCR, 7- to 12-mm single-shot precision (1–3 mm for a normal point) for a geodetic satellite, with overall accuracy at 8–18 mm, i.e. ∼1–2 cm. Normal points are made from a number of single shots, according to ILRS-prescribed guidelines; MOBLAS-6 averages about 66 shots (data points) per NP for LAGEOS. Limitations of space in this chapter preclude the discussion of all factors involved at great depth, but some of the issues involved in determining the range between the SLR station reference point and the satellite being tracked will become clear. The LAGEOS satellites will be used as example; principles involved will be more or less the same for other satellites.

An SLR station provides data in a specific format, as agreed to from time to time by the International Laser Ranging Service (ILRS) community (Pearlman et al. 2002). These data are uploaded to data centres and are consequently utilised to determine the range to a satellite; the ranges can then be used in a modelling process to estimate other parameters (Earth orientation, station position, gravity coefficients, etc.). The ranging data basically consist of information such as satellite identification number, system-specific details such as wavelength of laser (532 nm in case of MOBLAS-6), calibrated system delay (two way in picoseconds), pass RMS (picoseconds) and epoch of laser firing in 0.1-μs units. Included is the main observable which is the two-way time-of-flight corrected for system delay in

Table 9.1 Laser ranging error budget for the French SLR stations at the turn of the century. These values are still representative of most modern SLR stations. Adapted from Exertier et al. (2000)

Origin	Precision (mm)	Accuracy (mm)
Laser	4–5	
pulse	1	
width	4–5	
Detector	3–6	
start	1–3	
return	3–5	
Timer	2–3	
Clock	1–2	
Calibration	1	2–6
geometry		1–2
electronic		1–4
Depend. (Az, El)	1–3	
Instrument	6–9	2–6
Atmosphere	3–5	5–8
pressure		1–2
temperature		1
humidity		4–5
Target signature		
LAGEOS (COM, etc.)	1–3	1–3
Single shot	7–12	
Normal point	1–3	8–18

picoseconds. Other data are required to model the atmosphere and are also recorded, such as surface pressure, temperature and humidity. Raw ranges (taken at the transmit rate of the SLR, 5 Hz in the case of MOBLAS-6) are compressed to form a normal point.

The normal point (NP) data at a given epoch is taken here as the point of departure. This NP is converted easily to a normal point range in metres using the equation

$$\mathrm{NPR}_i = \left(\frac{\mathrm{NPtof}_i}{1 \times 10^{12}} \times c \right) \Big/ 2 \; (m), \tag{4}$$

where NPtof_i is the normal point time-of-flight (picoseconds) recorded at a certain epoch and c is (Kaplan 2005) the velocity of light (299,792,458.0 m/s). The range found in (4) needs to be corrected by taking into account the effects of the atmosphere (Δa_i in (5)), the centre-of-mass correction (CoM) of the satellite (0.251 m for LAGEOS 2), SLR station range bias and a relativistic correction. A range equation (5) can then be written as

$$NPR_i = \left(\frac{NPtof_i}{1 \times 10^{12}} \times c \right) \Big/ 2 - \Delta a_i + \Delta \mathrm{CoM}_i - \Delta R_{b_i} - \Delta GR_i - \Delta \varepsilon_i, \tag{5}$$

where NPR_i (m) is the normal point range, i.e. the observed range. The centre-of-mass correction is $\Delta\mathrm{CoM}_i$ and the range bias, general relativity correction and a correction for unknown random errors are $\Delta\mathrm{CoM}_i$, ΔR_{b_i} and $\Delta\varepsilon_i$, respectively. In Sect. 9.4 this observed range per normal point will be utilised to calculate the observed–calculated (O–C) residuals as part of the SLR data analysis process.

9.2.1 Atmospheric Delay Correction

At laser (optical) wavelengths the troposphere is dispersive and a correction for an additional delay due to the troposphere must be made. An often-used correction for atmospheric delay is given by Marini and Murray (1973). A modification of this approach is implemented by Mendes et al. (2002), which implements newly derived mapping functions for optical wavelengths, using a large database of ray tracing radiosonde profiles. These functions are optimised for a wavelength commonly used in SLR systems (532 nm) and are valid for elevation angles greater than 3°, if one neglects the contribution of horizontal refractivity gradients. Typical SLR data processing rejects data captured below a certain specified elevation, and software generally makes provision for a variable setting for rejection of data below any selected elevation. Rejecting data at low elevation will however reduce the geometrical strength of the measurements and increase the correlation between errors in the vertical coordinate and measurement biases.

At low elevations, this delay can be large (several metres), and therefore the model used for correcting the range should be very accurate. A current atmospheric delay model (Mendes and Pavlis 2004), adopted in October 2006 by the Analysis Working Group of the ILRS, which is also used in the SLR Data Satellite Analysis Software package developed at HartRAO (Combrinck and Suberlak 2007), makes provision for the different laser wavelengths used by current SLR equipment.

In order to demonstrate the model by Mendes and Pavlis (2004), Fig. 9.3 is a plot of atmospheric delay as a function of elevation for a 7-day arc of 11 SLR stations. Laser wavelengths at 423, 532 and 846 nm and a range of relative humidity, pressure and temperatures at ground level are represented, creating a band of corrections. Elevation cut-off was set at 15°, at which corrections are more than 8 m. The importance of accurate atmospheric modelling cannot be overstated and is an active area of research to improve accuracies. Two-colour lasers could be used to determine the delay in a similar way as dual frequency GNSS, DORIS and VLBI is used, but this is still in an experimental stage.

The atmospheric delay in the zenith direction (McCarthy and Petit 2003, 22 June 2007 update) experienced by the laser can be described by

$$d^z_{\mathrm{atm}} = d^z_h + d^z_{nh} = 10^{-6}\int_{r_s}^{r_a} N_h\, dz + 10^{-6}\int_{r_s}^{r_a} N_{nh}\, dz, \tag{6}$$

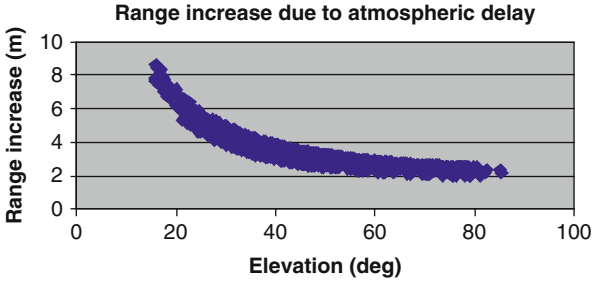


Fig. 9.3 The increase in range due to atmospheric delay, based on Mendes et al. (2002) and Mendes and Pavlis (2004), is the largest correction in the range model

when divided into hydrostatic $\left(d_h^z\right)$ and non-hydrostatic $\left(d_{nh}^z\right)$ components. In (6) the total group refractivity of moist air is denoted by $N = (n-1) \times 10^6$, n is the total refractive index of moist air, the hydrostatic (termed dry as it results from refractivity of dry gases in the troposphere, although it contains the non-dipole component water vapour refractivity) and non-hydrostatic (wet) components of the refractivity are given by $\left(d_h^z\right)$ and $\left(d_{nh}^z\right)$, respectively. The geocentric radius of the laser station is r_s, while r_a is the geocentric radius of the top of the neutral atmosphere; d_{atm}^z and dz have length units.

Mendes and Pavlis (2004) provide closed-form expressions to enable calculation of the zenith delay. The zenith hydrostatic component (in metres) is given by

$$d_h^z = 0.002416579 \frac{f_h(\lambda)}{f_s(\phi, H)} P_{\text{s}}, \tag{7}$$

where P_{s} is the surface barometric pressure (normally accurate to 0.1 hPa). The dispersion equation $f_h(\lambda)$ in (7) for the hydrostatic component is provided as

$$f_h(\lambda) = 10^{-2} \times \left[k_1^* \frac{\left(k_0 + \sigma^2\right)}{\left(k_0 - \sigma^2\right)^2} + k_3^* \frac{\left(k_2 + \sigma^2\right)}{\left(k_2 - \sigma^2\right)^2} \right] C_{\text{co}_2}, \tag{8}$$

where in (8), $k_0 = 238.0185\ \mu\text{m}^{-2}$, $k_2 = 57.362\ \mu\text{m}^{-2}$, $k_1^* = 19,990.975\ \mu\text{m}^{-2}$ and $k_3^* = 579.55174\ \mu\text{m}^{-2}$.
The wave number $\sigma = \lambda^{-1}$, with the wavelength λ in micrometres.

Following IAG recommendations, a CO_2 content (x_c) of 375 parts per million (ppm) should be used, so that parameter C_{co_2} in (8) is given by

$$C_{\text{co}_2} = 1 + 0.534 \times 10^{-6} (x_c - 450) = 0.99995995. \tag{9}$$

The denominator function of (7) for the zenith hydrostatic component is given by

$$f_{\text{s}}(\phi, H) = 1 - 0.00266 \cos 2\phi - 0.00000028H, \tag{10}$$

with ϕ being the geodetic latitude of the SLR station and H its geodetic height in metres. Considering the non-hydrostatic component d_{nh}^{z}, in metres, the equation

$$d_{nh}^{z} = 10^{-4}\left(5.316\, f_{nh}(\lambda) - 3.759\, f_{h}(\lambda)\right)\frac{e_{\mathrm{s}}}{f_{\mathrm{s}}(\phi, H)} \tag{11}$$

is given where e_s is the water vapour pressure at the surface, which can be calculated from the relative humidity and temperature (in degree Celcius) as

$$e_s = \frac{\left[6.1078 \times \left(\frac{7.5 \times t}{237.3 + t}\right)^{10}\right] \times RH}{100}. \tag{12}$$

Equation (12) includes relative humidity RH as a percentage. In (11) the dispersion formula for the non-hydrostatic component can be determined by

$$f_{nh}(\lambda) = 0.003101\left(\omega_0 + 3\omega_1\sigma^2 + 5\omega_2\sigma^4 + 7\omega_3\sigma^6\right), \tag{13}$$

with $\omega_0 = 295.235$, $\omega_1 = 2.6422\ \mu\mathrm{m}^2$, $\omega_2 = -0.032380\ \mu\mathrm{m}^4$and $\omega_3 = 0.004028\ \mu\mathrm{m}^6$. Once the total atmospheric delay in the zenith direction has been calculated, it needs to be mapped to the elevation at which the laser beam is being fired towards the satellite which is being tracked.

As the contribution to atmospheric refraction by water vapour is relatively small (at laser wavelengths) compared to the total refraction, a single mapping function can be used for SLR purposes (McCarthy and Petit 2003), so that the atmospheric delay in the direction of ranging d_{atm} is given by

$$d_{\mathrm{atm}} = d_{\mathrm{atm}}^{\mathrm{z}} \cdot m(e), \tag{14}$$

where d_{atm}^{z} is the total zenith propagation delay and $m(e)$ is the total mapping function (MF). The MF by Mendes et al. (2002) is based on a truncated form of the Marini (1972) continued fraction in terms of $1/\sin(e)$, normalised to unity. Until quite recently (at least until 2005), the Marini and Murray (1973) formulation for atmospheric delay was the IERS recommended standard for SLR. Mendes et al. (2002) provide the formulation where $m(e)$ is given as

$$m(e) = \frac{1 + \dfrac{a_1}{1 + \dfrac{a_2}{1 + a_3}}}{\sin e + \dfrac{a_1}{\sin e + \dfrac{a_2}{\sin e + a_3}}}. \tag{15}$$

Incorporating the values in Table 9.2 for the coefficients $a_i\ (i = 1, 2, 3)$, using the formulation for a_i as

$$a_i = a_{i0} + a_{i1}t_s + a_{i2}\cos\phi + a_{i3}H, \tag{16}$$

Table 9.2 Coefficients for FCULa (Mendes et al. 2002)

a_{ij}	**Coefficients**
a_{10}	0.00121008
a_{11}	$1.7295e^{-6}$
a_{12}	$3.191e^{-5}$
a_{13}	$-1.8478e^{-8}$
a_{20}	0.00304965
a_{21}	$2.346e^{-6}$
a_{22}	$-1.035e^{-4}$
a_{23}	$-1.856e^{-8}$
a_{30}	0.068777
a_{31}	$1.972e^{-5}$
a_{32}	–0.003458
a_{33}	$1.06e^{-7}$

where t_s is the temperature logged at the SLR station in degree Celcius (contained in normal point data file) and H is the station height in metres, it is then possible to make the atmospheric delay correction for the SLR range. As example, utilising the delay model described in this section, at a laser wavelength of 532 nm, relative humidity of 40%, pressure of 983.8 mb, a temperature of 29.41°C, an elevation change from 45.1 to 45.2° creates a decrease in range delay correction from 3.363 m to 3.357 m, i.e. a total of 6 mm with a 0.1° step in elevation.

9.2.2 Centre-of-Mass Correction

When the laser pulse reflects off the front face of the satellite, e.g. LAGEOS, there is still some distance from the face to the centre-of-mass point. The point of reflection is not clearly defined either, so that the reflected pulse is a convolution of the returns from the individual CCRs, each of which exhibits a different cross-section. The laser pulse therefore arrives at slightly different times (related to the planar wavefront first reaching the nearest CCR) at each individual CCR, which causes a broadening (about twice the 200-ps pulse length of MOBLAS-6) of the received pulse. In addition, a random phase delay is introduced by the differential delay between the individual CCRs, which could cause target "speckle". Such a speckle effect will occur if the temporal profiles from the CCRs overlap when received by the SLR detector, causing a random interference pattern (and random intensity) between the electric fields. This delayed return process is described very well in Degnan (1993), where it is assumed that, on average, the return waveform behaves as if the CCR is an incoherent source, so that the interference effect is ignored. Here we will discuss only the essentials required to understand the centre-of-mass correction required for our range model, using Fig. 9.4 as point of departure.

Following Degnan (1993), Fig. 9.4 describes some quantities necessary to discuss satellite impulse response; again we assume LAGEOS as an example of a typical geodetic satellite. A CCR's reflection centre is given by Arnold (1978) as

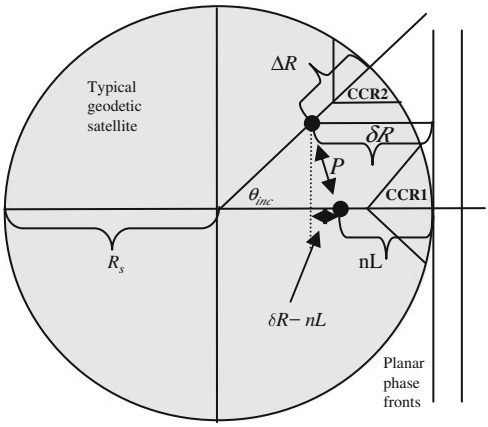


Fig. 9.4 Sketch of typical geodetic satellite such as LAGEOS, defining parameters required to discuss satellite impulse response. Adapted from Degnan (1993)

$$\Delta R\left(\theta_{\text{inc}}\right) = nL\sqrt{1-\left(\frac{\sin\theta_{\text{inc}}}{n}\right)^2} = nL\cos\theta_{\text{ref}}, \tag{17}$$

where $\Delta R\left(\theta_{\text{inc}}\right)$ is taken as the distance between the centre of the CCR's front face to the reflection point P_r, the vertex of the CCR to its front face distance is denoted by L, the refractive index of the CCR material is n, the angle of incidence is θ_{inc} and the corresponding refraction angle is θ_{ref}. The reason for the broadening of the received pulse is clearly illustrated in Fig. 9.4 by the distance $\delta R - nL$, which can be ascribed to the fact that CCR2 is located at an incident angle θ_{inc} to the laser pulse, which results in a time delay (Degnan 1993) $\Delta t\left(\theta_{\text{inc}}\right)$, where

$$\begin{aligned}\Delta t\left(\theta_{\text{inc}}\right) &= \frac{2}{c}\left\{R_s - \left[R_s - \Delta R\left(\theta_{\text{inc}}\right)\right]\cos\theta_{\text{inc}}\right\} \\ &= \frac{2R_s}{c}\left\{1-\cos\theta_{\text{inc}}\left[1-\frac{nL}{R_s}\cos\theta_{\text{ref}}\right]\right\}\end{aligned} \tag{18}$$

compared to CCR1 where $\theta_{\text{inc}} = 0$. In (18), with reference to Fig. 9.4, R_s is the radius of the satellite (0.298 m in the case of LAGEOS). Utilising this simple model it is then possible to determine an average (basically the convoluted response on a short laser pulse) satellite temporal response through the addition of the returns from each of the CCRs.

This is described in detail by Degnan (1993) and here we will therefore continue to the result, which finds that if a centre-of-mass correction is calculated from the centre of the impulse response profile, a value 250.2 mm is obtained for LAGEOS.

This is in good agreement with the current ILRS accepted value for a wavelength of 532 nm, which is 251 mm as determined by Otsubo and Appleby (2003). The ILRS maintains a web page which lists appropriate corrections for different SLR

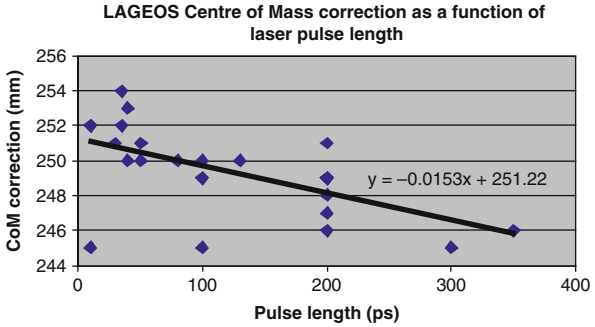


Fig. 9.5 ILRS recommended centre-of-mass (CoM) corrections for LAGEOS vs. laser pulse length

stations; these data can be incorporated as a set-up file for analysis software. The value quoted for MOBLAS-6 is 247 mm.

The 251 mm is an average value, as the returned signal will always be a convolution of the individual returns from the satellite, and slight deviations of ±2 mm will occur during actual ranging as LAGEOS rotates and therefore changes the incident angles θ_{inc} of the laser pulse. The deviation from 251 mm could be affected by station-specific parameters such as type of detector, laser frequency and pulse length and transmit power.

In Fig. 9.5 the data for CoM corrections as published on the ILRS website are plotted as a function of laser pulse length, capturing the 251 mm as a constant in the equation of the line. In general, a shorter pulse length requires a larger CoM correction, with the exception of Herstmonceux when operating at 1 kHz and 10-ps pulse length and when operating at Hertz level and 100-ps pulse length, causing two outliers and distortion of the function. The complexity of CoM is illustrated in Fig. 9.5; many different factors contribute to making up the CoM correction for a specific station. This plot is generated from LAGEOS_CoM_Table_081023.pdf, source: http://ilrs.gsfc.nasa.gov/docs/.

9.2.3 SLR Station Range and Time Bias

An additional correction to the range which has to be incorporated is the station range bias. This station range bias is a bit open to interpretation and different implementation. Most of the time, it would be more correct to use the term *computationally determined range bias* and in fact does not always refer to a real station-dependent error.

On the long term, averaged over several months, independent range biases from different processing centres should agree at some level. Short arc range bias, however, could be problematic due to the correlation between range bias and station position, especially the vertical coordinate. A positive range bias (say +5 mm) means

the range measured is too long and should be subtracted. A time bias of 5μs means the epoch of registering the observation is 5μs late and should be subtracted from the NP time tag.

Heuristically speaking, range bias should be a station-dependent error; therefore it should be due to some non-linearities in the interval counter (creating a non-linear bias) or errors in the barometric pressure, temperature and relative humidity sensors (these last three parameters affecting atmospheric delay modelling). The determined station position (ITRF point), an error in system delay, an error in the tie (eccentricity) between the station coordinate reference point and the SLR telescope reference point could be added to the range error budget.

Unstable calibration targets can produce induced range biases between passes, as the apparent calibrated "system delay" determined by pre- and post-system calibrations will be affected. Typical rms values for MOBLAS-6 obtained while calibrating the SLR system using reference piers equipped with CCRs are ∼0.4 cm. Husson (1993) reports other factors as well, such as insertion of neutral density (ND) filters, which could induce systematic biases of 1–2 mm.

Another example quoted by Husson (1993) is range bias of several millimetres due to signal strength issues. This effect is due to the fact that the length of the pulse is related to the range in that, when the return signal is weak, the probability of detection is a function of the total pulse width. This signal strength variation-induced bias varied from pass to pass and had slight elevation dependence, low-elevation data being biased longer by some millimetres than higher elevation data. The effect of signal strengths was reduced after calibration techniques were improved and a new photo multiplier was installed.

The telescope reference point is determined by the intersection of the telescope axes (if *XY* mount) or the axis offset and its right-angle projection from the elevation axis onto the azimuth axis (if an Az–El mount). Site and system calibration should find any systematic or other errors to millimetre accuracy. The SLR eccentricity is supposed to be accurate within 1 mm; however, this is very difficult to achieve in practice. One can therefore assume that most, if not all, SLR stations have a small bias between the fixed to bedrock reference marker (the ITRF reference point) and the virtual SLR system reference point (eccentricity). Nevertheless, one also assumes that all effort was undertaken to determine these different distances and offsets as accurately as possible.

When processing the normal point data from a station, the station reference point in the ITRF (at a certain epoch, say ITRF2005) plus an adjustment for station velocity due to tectonic plate motion is taken as the observation epoch. This adjustment is approximated by

$$\left.\begin{aligned} X_{\mathrm{oe}} &= X_I + ((O_{\mathrm{MJD}} - \mathrm{ITRF}_{\mathrm{MJD}})/365.25) \times \Delta X_{\mathrm{yr}}, \\ Y_{\mathrm{oe}} &= Y_I + ((O_{\mathrm{MJD}} - \mathrm{ITRF}_{\mathrm{MJD}})/365.25) \times \Delta Y_{\mathrm{yr}}, \\ Z_{\mathrm{oe}} &= Z_I + ((O_{\mathrm{MJD}} - \mathrm{ITRF}_{\mathrm{MJD}})/365.25) \times \Delta Z_{\mathrm{yr}}. \end{aligned}\right| \tag{19}$$

In (19) X_{oe}, Y_{oe}, Z_{oe} are the station reference coordinates at observation epoch, X_I, Y_I, Z_I are the station reference coordinates in the ITRF at the particular ITRF epoch, O_{MJD} is the epoch of observation, $ITRF_{MJD}$ is the ITRF epoch, (say both in modified Julian day for computational convenience) and $\Delta X_{yr}, \Delta Y_{yr}, \Delta Z_{yr}$ are the site velocities determined in the ITRF. This is a linear approximation and does not take into account rotation of the tectonic plate; this can be accommodated by including a factor for angular rotation. If processing centres utilise different ITRF epochs or velocities, range bias differences will result.

During processing, an attempt will be made to determine the small offset due to any one or combination of the above-mentioned errors. This can be done in various ways; for instance, it can be estimated as a solve-for parameter during a least-squares adjustment of the observed–computed residuals. It is very difficult to separate the range bias from station position related errors if the tracking data are not sampled uniformly across the sky, i.e. from horizon to horizon. Sometimes operators track a satellite when it comes into view, but switch over to another satellite before the satellite has moved past its zenith point.

As an example, if the satellite was moving towards the observer, and there is a negative range bias of say –5 mm, assuming the vector from observer to satellite is biased by a 5-mm positional error, the ranging would indicate first short, then too long after the satellite passed zenith. The ±5 mm could help indicate positional error. If the operator switched to another satellite or stopped tracking for a lunch break, only a short range would be recorded, as if there were a constant negative bias. The range bias is also related to the station time bias, which exists due to the small (microsecond) offset between true time (UTC) and station time (UTC + offset) and any uncalibrated instrumental phase delays.

Figure 9.6 illustrates the concepts of range and time bias. If a complete pass is available, enough information exists to determine slopes, fit slope parameters, etc.; otherwise the results are FT_b and FR_b, false time and range biases. The determination of range biases is also absorbed into the analysis process and will be different for alternative analysis software and even different for alternative a priori conditions for the same software. It is easy to see why; range bias is also dependent on the computed orbit, and the assumptions made to compute the propagation of this orbit; these assumptions include pre-set solve-for parameters, their a priori estimates and errors, gravity model used and a vast number of alternative processing strategies.

It is therefore no wonder that, for the same SLR stations, tracking the same satellite, i.e. for the same data, different and sometimes opposite sign values are determined by processing. Determining R_b utilising the NP tracking data is therefore not so straightforward. Apart from solving for it by adding it to the number of solve-for parameters, it can be done quite easily pass by pass or arc by arc by obtaining the average observed–computed (O–C) residuals per station. It can also be obtained by fitting high-order polynomials to the observed and calculated ranges and calculating the difference at minima. This value so obtained can be re-inserted into the least squares process or into whatever adjustment procedure is used. The time bias T_b can be obtained from the slopes of the observed and calculated ranges.

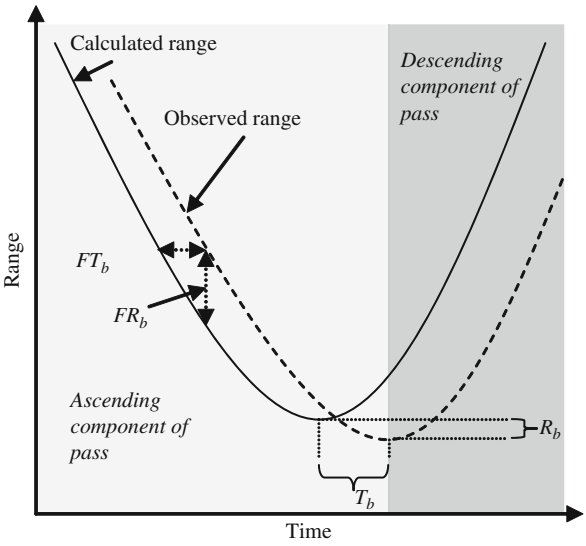


Fig. 9.6 The concept of range and time bias illustrated. It is clear that a complete pass, i.e. ascending and descending components of pass, is preferable when trying to estimate either a range or a time bias. Short passes lead to erroneous values

Care needs to be exercised, however, to avoid erroneous values being fed back into the solution. For instance, if less than 8 NPs have been collected for a pass, the HartRAO SDAS software does not estimate a time bias, as the polynomials that fit to the observed and computed ranges cannot capture the true slope with partial information. The recording and characterisation of the ILRS network biases is a necessary and arduous task; it is critical to the analysis of SLR data if millimetre accuracy is to be achieved. To this extent, the ILRS has developed a standard SINEX (Software INdependent EXchange) format for documenting biases.

9.2.4 Relativistic Range Correction

This is the last correction to the range to be discussed; it is at the level of range bias corrections and is about 7 mm for LAGEOS. The correction is described in IERS (2003) in detail. Following IERS (2003):

$$t_2 - t_1 = \frac{|\vec{x}_2(t_2) - \vec{x}_1(t_1)|}{c} + \sum_J \frac{2GM_J}{c^3} \ln\left(\frac{r_{J1} + r_{J2} + \rho}{r_{J1} + r_{J2} - \rho}\right), \qquad (20)$$

where $t_2 - t_1$ is the total time delay when a laser pulse at coordinate x_1 is transmitted at time t_1 and the return pulse is received at coordinate x_2 at time t_2. In (20), $r_{J1} = \left|\overrightarrow{x_1} - \overrightarrow{x_J}\right|$, $r_{J2} = \left|\overrightarrow{x_2} - \overrightarrow{x_J}\right|$ and $\rho = \left|\overrightarrow{x_2} - \overrightarrow{x_1}\right|$ is the range. In this formulation the sum is carried over all bodies J with mass M_J centred at x_J. Following Ries et al. (1988), in the case of SLR the only body to be considered for near-Earth satellites is the Earth, as analysis is done in the geocentric frame of reference. In

practice, ρ is the range between receiver and transmitter before the correction for relativity is applied, but after the other corrections have been applied.

9.3 Force and Orbital Model

There are many applications of SLR data, and most are dependent on ILRS products such as SLR station position and velocities in the International Terrestrial Reference Frame, Earth orientation parameters (length of day and polar motion), coefficients of Earth's gravity field and their variation in time and highly accurate (within centimetres) satellite ephemerides. Some of the applications are in fields other than space geodesy, which illustrate the diverse and multidisciplinary fundamental support to science provided by SLR. As an example, the current international scientific agenda to support long-term studies which relate to Earth system change (ocean levels, global climate, etc.) is supported by SLR products as it supports the monitoring of the Earth system by providing orbital calibration of remote sensing and other satellites.

The SLR products also support studies of the three-dimensional deformation of the solid Earth, measurement of the variation in Earth's liquid volume (ocean circulation, mean sea level, ice sheet thickness) amongst a list of applications. Other applications such as intercontinental time transfers, T2L2 (time transfer by laser link) using SLR, initiated by the Observatoire de la Côte d'Azur (France) are being developed (Vrancken et al. 2008) which will improve the synchronisation of remote clocks on Earth by at least an order of magnitude. This technique is based on the propagation of light pulses in space which is better characterised than propagation of radio waves.

9.3.1 Introduction

Several scientific software packages are in use by the scientific community to analyse SLR data and perform precise orbit determination (POD) and geodetic parameter estimation. Notable amongst these are GEODYN II (Pavlis et al. 1999), developed at NASA Goddard Space Flight Centre (GSFC), which is a benchmark software for POD, and a commercial version of GEODYN II, MicroCosm®, a very comprehensive high-precision orbit and geodetic parameter determination software system (http://www.vmsi-microcosm.com). Other software such as the SATellite ANalysis (SATAN) software (Sinclair and Appleby 1986) developed at Royal Greenwich Observatory, Herstmonceux, is being utilised for SLR analysis and POD.

These and other POD packages are constantly being developed and expanded to include new models and new applications. HartRAO has been developing an SLR POD and geodetic parameter estimation software package (named SLR Data

Analysis Software, SDAS), which uses SLR data through a development and extension of the spaceflight dynamics library provided by Montenbruck and Gill (2001) of the Deutsches Zentrum für Luft- und Raumfahrt (DLR), Germany.

9.3.2 Orbital Modelling

This chapter can only devote minimal space to force and orbital modelling, so this section is brief, and will concentrate more on the adjustments and compensations required that will minimise the O–C residuals than on orbital propagation methods and techniques. Nevertheless, it will hopefully provide a quick glance at some of the main issues involved.

9.3.3 Force Model

A satellite orbiting a spherical Earth undergoes acceleration as described by Newton's law of gravity:

$$\ddot{\vec{r}} = -\frac{GM_{\oplus}}{r^3}\vec{r}, \tag{21}$$

where $\ddot{\vec{r}}$ is the acceleration in a geocentric reference frame (the fraction $-\vec{r}/r$ in (21) is the unit vector from the satellite to the geocentre), $GM_{\oplus}$ is the product of Earth's gravitational constant and its mass, $\vec{r}$ is the position vector of the satellite and r is the geocentric range given by $\sqrt{(\vec{r}\cdot\vec{r})}$. Due to the fact that Earth is not spherical and that there are many different forces acting on the satellite and thereby affecting its orbit, one has to model these forces and account for them in the orbital calculations. Equation (21) can therefore be expanded to

$$\ddot{\vec{r}} = -\frac{GM_{\oplus}}{r^3}\vec{r} + \vec{f}, \tag{22}$$

where (following and extending the notation of Tapley et al. 2004) the perturbing force $\vec{f}$ is made up of a range of forces acting on the satellite:

$$\vec{f} = \vec{f}_{\mathrm{NS}} + \vec{f}_{\mathrm{TC}} + \vec{f}_{\mathrm{3B}} + \vec{f}_{g} + \vec{f}_{\mathrm{Drag}} + \vec{f}_{\mathrm{SRP}} + \vec{f}_{\mathrm{ERP}} + \vec{f}_{\mathrm{other}} + \vec{f}_{\mathrm{Emp}}. \tag{23}$$

In (23) $\vec{f}_{\mathrm{NS}}$ is the force resulting from the uneven mass distribution on the earth and is found from the gradient of the gravitational potential U; it also includes the variation of Earth's mass distribution due to ocean and earth-tides. In addition to the static gravity field, the contribution of the temporal variations of the static gravity field is included in $\vec{f}_{\mathrm{TC}}$. A three-body (n-body) perturbation caused by the gravitational forces from the Moon, Sun and planets are denoted by $\vec{f}_{3B}$, the effects of general relativity is accommodated by $\vec{f}_{g}$, $\vec{f}_{\mathrm{Drag}}$ is the drag caused by the atmosphere, $\vec{f}_{\mathrm{SRP}}$ is

the solar radiation force contribution, $\vec{f}_{\mathrm{ERP}}$ is the earth radiation pressure and $\vec{f}_{\mathrm{Other}}$ contains other forces such as thermal, satellite rotation-dependent effects. Empirical corrections, normally expressed in a local frame, split into radial, tangential and normal (RTN) components are denoted as $\vec{f}_{\mathrm{Emp}}$. In the following sections, a brief discussion of these perturbing forces is made; the literature contains adequate and in-depth discourses (cf. Hoffman-Wellenhof and Moritz 2005), so only essentials are discussed here following the notation of Tapley et al. (2004) for convenience.

9.3.3.1 Gravitational Perturbations $\left(\vec{f}_{\mathrm{NS}}\right)$

The gravitational potential between two point masses can be described by

$$U = \frac{GM_1M_2}{r}, \tag{24}$$

where r is the distance between the two masses. Using the gradient of U, one can obtain the gravitational force on M_2:

$$\vec{F} = \nabla U = -\frac{GM_1M_2\vec{r}}{r^3}, \tag{25}$$

where $\vec{r} = x\vec{u}_x + y\vec{u}_y + z\vec{u}_z$ is the position vector of M_2 in relation to M_1, and $\nabla = \frac{\partial}{\partial x}\vec{u}_x + \frac{\partial}{\partial y}\vec{u}_y + \frac{\partial}{\partial z}\vec{u}_z$. Here $\vec{u}_x, y\vec{u}_y$ and $z\vec{u}_z$ are unit vectors.

Considering the gravitational potential U(Tapley et al. 2004) affecting a point mass, m', at a location external to a body M, with M having an arbitrary mass distribution and being modelled as a collection of point masses, one can write

$$U = m' \iiint \frac{G\gamma \, dx\, dy\, dz}{\rho}, \tag{26}$$

with γ representing the mass density of mass element dm, the differential volume is $dx\, dy\, dz$, ρ is the distance between dm, and m'. The potential as per (26) can be written in compact form if the external mass is taken as unity so that $m' = 1$ and the integral notation is taken to be over the complete mass of the body so that

$$U = \int_M \frac{G\ dm}{\rho}. \tag{27}$$

Equation (27) can then be written as an infinite series:

$$U = \frac{G}{r} \int_M \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^l P_l\,(\cos S)\, dm, \tag{28}$$

with the distance between the origin of the total mass M and the differential mass dm denoted by R. The Legendre polynomial of degree l is P_l. The argument of P_l is

$\cos S$, where S is the angle between the vector $\vec{R}$(origin to dm) and $\vec{r}$(origin to m'). In order to evaluate the integral of (28), the Legendre polynomial can be expanded into spherical harmonics:

$$\begin{aligned} U &= \frac{\mu}{r} + U', \\ U' &= -\frac{\mu^*}{r} \sum_{l=1}^{\infty} \left(\frac{a_e}{r}\right)^l P_l(\sin\phi)\, J_l \\ &\quad + \frac{\mu^*}{r} \sum_{l=1}^{\infty} \sum_{m=1}^{l} \left(\frac{a_e}{r}\right)^l P_{lm}(\sin\phi)\, [C_{lm}\cos m\lambda + S_{lm}\sin m\lambda]\,. \end{aligned} \tag{29}$$

In (29), m' has spherical coordinates (r, ϕ, λ) , which can be written as

$$\begin{aligned} x &= r\cos\phi\cos\lambda, \\ y &= r\cos\phi\sin\lambda, \\ z &= r\sin\phi, \end{aligned} \tag{30}$$

where the (x, y, z) system is fixed in the body at its origin O. In (29), a reference distance a_e and reference mass ($\mu^* = GM^*$) are included (M^* is the mass of the Earth in our case) as scale factors to non-dimensionalise C_{lm} and S_{lm}, the mass coefficients. Legendre's associated functions are denoted by P_{lm} of degree and order l and m , respectively. The spherical harmonic coefficients J_l, C_{lm} and S_{lm} represent mass properties of the body. Coefficients that describe the part of the potential that does not depend on longitude are called zonal coefficients, and these are related to $C_{l,m}$ through the relation $J_l = -C_{l,0}$. Tesseral harmonics refer to C_{lm} and S_{lm} with $l \neq m$ and the coefficients are termed sectorial harmonics if $l = m$. The geopotential coefficients vary over a large range of ten or more orders of magnitude. In order to cope with this, the coefficients can be normalised, so that they are more of the same magnitude. Gravity models commonly used for SLR such as the GRACE model GGM02 (Tapley et al. 2005) are published in normalised format. The normalised coefficients are defined as (Montenbruck and Gill 2001)

$$\begin{Bmatrix} \bar{C}_{lm} \\ \bar{S}_{lm} \end{Bmatrix} = \sqrt{\frac{(l+m)!}{(2-\delta_{0m})\,(2l+1)\,(l-m)!}} \begin{Bmatrix} C_{lm} \\ S_{lm} \end{Bmatrix}. \tag{31}$$

If normalised coefficients are used, a normalised set of Legendre functions must be used. In the HartRAO SDAS software, the normalised coefficients are converted back to unnormalised coefficients; modern computers and compilers do not have a problem in processing the large range of numbers. Since the degree one terms (J_1, C_{11} and S_{11}) are directly related to the offset from the origin O to the centre of mass of the body, in the case of geocentric coordinate systems such as commonly used in SLR POD, the degree one terms are zero. When the z axis is aligned with the maximum moment of inertia, then $C_{21} = 0$ and $S_{21} = 0$. If we assume that m' represents a satellite one can write (Tapley et al. 2004)

$$\ddot{\vec{r}} = \nabla U = -\frac{GM_{\oplus}}{r^3}\vec{r} + \vec{f}_{\mathrm{NS}}, \tag{32}$$

as m'/M is very small. In (32) the force contribution by non-spherical terms are presented by f_{NS}, i.e. $\nabla U'$. If the acceleration term is expressed in body-fixed and the gravitational potential in spherical coordinates we have

$$\nabla U = \frac{\partial U}{\partial r}\vec{u}_r + \frac{1}{r}\frac{\partial U}{\partial \phi}\vec{u}_\phi + \frac{1}{r\cos\phi}\frac{\partial U}{\partial \lambda}\vec{u}_\lambda, \tag{33}$$

where the gradient gives force components in spherical coordinates. These can be rotated via a coordinate transformation into (x, y, z) components using

$$T^{xyz}_{r\phi\lambda} = \begin{bmatrix} \cos\phi\cos\lambda & -\sin\phi\cos\lambda & -\sin\lambda \\ \cos\phi\sin\lambda & -\sin\phi\sin\lambda & \cos\lambda \\ \sin\phi & \cos\phi & 0 \end{bmatrix}. \tag{34}$$

If one wants to provide $\ddot{\vec{r}}$ in a non-rotating system, then an additional transformation is required, e.g. if the two axes Z and z coincide and need to be rotated through an angle α, the required transformation is

$$T^{XYZ}_{xyz} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{35}$$

Therefore, the portion of the perturbing force contributed by the mass distribution of Earth is written as (Tapley et al. 2004)

$$\vec{f}_{NS} = T^{XYZ}_{xyz} T^{xyz}_{r\phi\lambda} \nabla U', \tag{36}$$

if $\vec{r}$ is in the non-rotating system (X, Y, Z) . In practical terms, as most calculations during POD estimation are done in an inertial reference system, complex transformations need to be made to transform from the Earth-fixed geocentric system as described here to the J2000 Earth-centred-inertial (ECI) system.

If T^{xyz}_{XYZ} is the transformation matrix from J2000 to Earth-fixed, then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathrm{ECF}} = T^{xyz}_{XYZ} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\mathrm{ECI}}, \tag{37}$$

where the transformation matrix can be expanded into

$$T^{xyz}_{XYZ} = WS'NP, \tag{38}$$

where P applies precession from a specified epoch to current time, N applies nutation to the current time, S' applies the rotation to allow for sidereal time and W

applies polar motion to align the true pole (z axis) with the pole of the ECF system. These matrices are described in detail in Tapley et al. (2004), Montenbruck and Gill (2001), Kaplan (2005) and McCarthy and Petit (2003).

9.3.3.2 Temporal Changes of the Gravity Field $\left(\vec{f}_{\mathrm{TC}}\right)$

Gravitational forces from the Sun and Moon result in a time-varying change of the shape of Earth; changes to the solid body are termed Earth-tides and the response of oceans due to lunisolar tidal perturbations we know as ocean-tides. In addition, the contributions of the solid Earth pole-tide and ocean pole-tide can be considered, which we will not discuss here; a description can be found in McCarthy and Petit (2003). These shape deformations of the earth modulate the static gravity field.

Changes resulting from solid Earth-tides can be modelled as variations in the standard geopotential coefficients C_{lm} and S_{lm}(Eanes et al. 1983; Montenbruck and Gill 2001; McCarthy and Petit 2003). Corrections to the unnormalised geopotential coefficients can be calculated using the formulation (Sanchez 1974; Montenbruck and Gill 2001)

$$\left\{\begin{array}{c}\Delta C_{lm}\\ \Delta S_{lm}\end{array}\right\} = 4k_l\left(\frac{GM}{GM_\oplus}\right)\left(\frac{R_\oplus}{s}\right)^{(l-1)}\sqrt{\frac{(l+2)\,(l-m)!^3}{(l+m)!^3}}P_{lm}\left(\sin\theta\right)\left\{\begin{array}{c}\cos\left(m\lambda\right)\\ \sin\left(m\lambda\right)\end{array}\right\}, \tag{39}$$

where k_l are the Love numbers for a given degree l. The distance to the tide generating body is s, $R_\oplus$is the geocentric radius of the Earth (6,378.1366 km), the Earth-fixed latitude and longitude of the disturbing body is ϕ and λ, respectively. A detailed description of the computations required can be found in McCarthy and Petit (2003).

In addition to the solid Earth-tides, ocean-tides need to be accounted for; both these effects on the static gravity field are complex and suffice it to say that the ocean-tide contribution can be incorporated in the normalised coefficients by spherical harmonic expansion of an ocean-tide potential (Eanes et al. 1983; Montenbruck and Gill 2001; McCarthy and Petit 2003):

$$\left\{\begin{array}{c}\Delta C_{lm}\\ \Delta S_{lm}\end{array}\right\} = \frac{4\pi GR_\oplus^2\rho_{\mathrm{w}}}{GM_\oplus}\frac{1+k'_l}{2l+1}\left\{\begin{array}{l}\sum\limits_{s(l,m)}\left(C^+_{\mathrm{slm}}+C^-_{\mathrm{slm}}\right)\cos\theta_s+\left(S^+_{\mathrm{slm}}+S^-_{\mathrm{slm}}\right)\sin\theta_s\\ \sum\limits_{s(l,m)}\left(S^+_{\mathrm{slm}}+S^-_{\mathrm{slm}}\right)\cos\theta_s+\left(C^+_{\mathrm{slm}}+C^-_{\mathrm{slm}}\right)\sin\theta_s\end{array}\right\}, \tag{40}$$

where k'_l $\left(k'_2=-0.3075,\ k'_3=-0.195, k'_4=-0.132, k'_5=-0.1032, k'_6=-0.0892\right)$ are the load deformation coefficients and ρ_{w} is the density of seawater $\left(1,025\ \mathrm{kg\,m^{-3}}\right)$. Ocean-tide coefficients (m) for the tide constituent s are denoted by $C^{\pm}_{\mathrm{slm}}$ and $S^{\pm}_{\mathrm{slm}}$. The argument of the tide constituent s as defined in the solid tide model presented in McCarthy and Petit (2003) is denoted by θ_s, the weighted summation of the six Doodson variables which are closely related to the arguments of the nutation theory (cf. (5) of McCarthy and Petit 2003).

9.3.3.3 Three-Body Perturbing Acceleration $\left(\vec{f}_{3B}\right)$

The satellite being tracked by the SLR station undergoes acceleration caused by the gravity of the Sun, Moon and planets. Considering the three-body perturbing force contribution to (22), where the equation of motion was described, one can write (Tapley et al. 2004), assuming that the additional bodies are point masses:

$$\vec{f}_{3B} = \sum_{j=1}^{n_p} GM_j \left(\frac{\vec{\Delta}_j}{\Delta_j^3} - \frac{\vec{r}_j}{r_j^3} \right), \tag{41}$$

where the summation is from one to the number of additional bodies (Sun, Moon, planets). In (41) a specific body is denoted by j, the gravitational parameter of each body j is represented by GM_j , $\vec{\Delta}_j$ is the position vector of the body j relative to the satellite, $\vec{r}_j$ is the position vector of the body j relative to Earth.

In practice, the position vectors of the additional bodies are found using precise ephemerides in the form of Chebyshev approximations such as JPL DE405 (Standish 1998), which is also utilised in the HartRAO SLR analysis software. This ephemeris is based on the International Celestial Reference Frame (ICRF) and is within 0.01 arc-seconds of the frame (cf. http://ssd.jpl.nasa.gov/).

9.3.3.4 General Relativity Contribution to the Perturbing Force $\left(\vec{f}_{\mathrm{g}}\right)$

The IERS 2003 (McCarthy and Petit 2003) recommendations as on 15 October 2004 discuss the relativistic correction to the acceleration of a satellite in Earth orbit where

$$\begin{aligned} \Delta\ddot{\vec{r}} = & \frac{GM_\oplus}{c^2 r^3} \left\{ \left[2(\beta+\gamma) \frac{GM_\oplus}{r} - \gamma \left(\dot{\vec{r}} \cdot \dot{\vec{r}} \right) \right] \vec{r} + 2(1+\gamma) \left(\vec{r} \cdot \dot{\vec{r}} \right) \dot{\vec{r}} \right\} + \\ & (1+\gamma) \frac{GM_\oplus}{c^2 r^3} \left[\frac{3}{r^2} \left(\vec{r} \times \dot{\vec{r}} \right) \left(\vec{r} \cdot \vec{J} \right) + \left(\dot{\vec{r}} \times \vec{J} \right) \right] + \\ & \left\{ (1+2\gamma) \left[\dot{\vec{R}} \times \left(\frac{-GM_s \vec{R}}{c^2 R^3} \right) \right] \times \dot{\vec{r}} \right\} \end{aligned} \tag{42}$$

is the correction which includes as

- first term, the non-linear Schwarzschild field of the Earth $\left(\approx 10^{-9}\ \mathrm{m\,s^{-2}}\right)$
- second term, Lense–Thirring precession (frame dragging) $\left(\approx 10^{-11}\ \mathrm{m\,s^{-2}}\right)$
- third term, de Sitter (geodesic) precession $\left(\approx 10^{-11}\ \mathrm{m\,s^{-2}}\right)$

Here the approximate magnitude of acceleration refers to LAGEOS as calculated by the HartRAO SDAS software. In (42), c is the speed of light, β, γ are the PPN parameters which equal 1 in general relativity, $\vec{r}$ is the position of the satellite relative to the Earth, $\vec{R}$ is the position of the Earth relative to the Sun, $\vec{J}$ is the Earth's angular momentum per unit mass ($\left|\vec{J}\right| \cong 9.8 \times 10^8\ \mathrm{m^2\,s^{-1}}$), $GM_\oplus$ is the gravitational coefficient of Earth and GM_s the gravitational coefficient of the Sun. These

effects are small, but cannot be ignored for POD purposes as there are some long-term periodic and secular effects (Huang and Liu 1992) of the orbit. In particular, the Schwarzschild effect causes perigee precession; de Sitter precession can lead to long-period variations of some of the orbital elements (Ω, ω, M) and Lense–Thirring precession causes secular rates (Ciufolini and Wheeler 1995) in the orbital elements Ω and ω.

9.3.3.5 Atmospheric Drag $\left(\vec{f}_{\mathbf{Drag}}\right)$

Atmospheric drag decelerates a satellite, especially low-altitude satellites. Modelling of the forces resulting from the satellite moving through the atmosphere is difficult for several reasons. Upper atmosphere densities are not well known, the interaction of neutral gas and charged particles with the satellite surfaces is complex and the change in orientation of the satellite in relation to the atmospheric particle flux has to be considered. For an in-depth discussion see Montenbruck and Gill (2001). Here we present the basic equation which can be used to approximate the negative acceleration experienced by the satellite. In (23),

$$\vec{f}_{\text{Drag}} = -\frac{1}{2} C_D \frac{A}{m} \rho v_r^2 \vec{e}_v, \tag{43}$$

where C_D is a dimensionless quantity that describes the satellite's interaction with the atmosphere. Typical values (Montenbruck and Gill 2001) for the drag coefficient ranges from 1.5 to 3.0. In (43), m is the mass of the satellite and the direction of the acceleration due to drag is parallel but opposite to the satellite velocity vector as described by the unit vector $\vec{e}_v = \vec{v}_r / v_r$. The atmospheric density is denoted by ρ. For satellites at the height of LAGEOS (~6,000 km), the effect of atmospheric drag is negligible, but needs to be taken into account if very high accuracies are required, e.g. for the estimation of relativistic effects.

9.3.3.6 Solar Radiation Pressure $\left(\vec{f}_{\mathbf{SRP}}\right)$

When a satellite is exposed to solar radiation, it absorbs and reflects photons; this imparts a force $\vec{f}_{\text{SRP}}$ on the satellite causing acceleration which is dependent on the solar flux, the satellite's mass m and cross-section A. Many complex models are used for specific satellites, but for the purpose of this chapter we assume that the satellite has an uncomplicated shape (no parabolic reflectors, solar panels, etc.) and that the surface normal points to the Sun. With reference to (23), one can then write (Vallado 2001; Montenbruck and Gill 2001) the solar pressure force contribution to the total perturbive acceleration as

$$\ddot{\vec{r}} = -\nu P_e C_R \frac{A}{m} \frac{\vec{r}_e}{r_e^3} AU^2. \tag{44}$$

In (44), P_e is the radiation flux from the Sun $\left(\approx 4.56 \times 10^{-6}\,\mathrm{N\,m^{-2}}\right)$, $\vec{r}_e$ is the geocentric position of the Sun and ν is a shadow function to cope with partial or total eclipses when the satellite moves into shadow. The conditions for the shadow functions are $\nu = 0$ if the satellite is in the shadow region (umbra), $\nu = 1$ if the satellite is in full sunlight and $0 < \nu < 1$ if the satellite is in partial shadow (penumbra). As example, an accepted value for LAGEOS is $C_R = 1.13$. This value can be used as a priori start-up value for SLR data analysis. The reflection coefficient can be estimated as a free parameter, which allows estimation of C_R to some degree of accuracy and without a priori knowledge of satellite orientation and reflectivity.

9.3.3.7 Earth Radiation Pressure $\left(\vec{f}_{\mathrm{ERP}}\right)$

An additional force $\left(\vec{f}_{\mathrm{ERP}}\right)$ imparted to the satellite is due to the reflection and scattering of solar radiation incident on the Earth's surface back onto the surface of the satellite. The amount of reflected light as a fraction of the incident shortwave solar radiation (see Fig. 9.7) is termed the albedo factor, a, which for the Earth has an average value of 0.34. Rubincam and Weiss (1986) list a number of factors which make modelling of the reflected solar pressure difficult:

- The irradiance on the satellite is dependent on the distance as well as on the orientation of the reflecting surface element (of the Earth).
- The irradiance depends on the solar zenith angle as well.
- The force is made up only of contributions from the elements illuminated by the Sun and, of these, only those visible by the satellite.

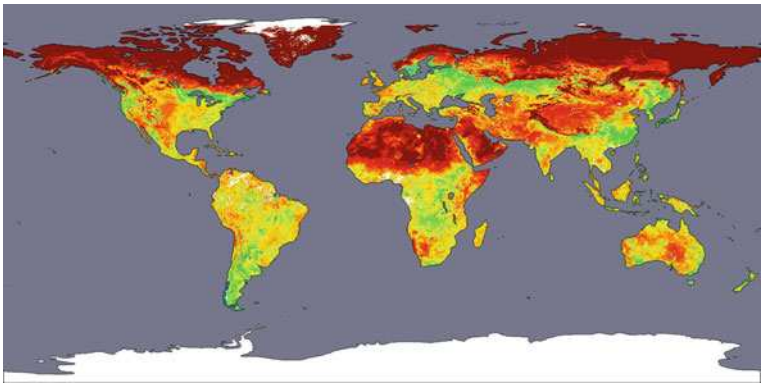


Fig. 9.7 Global albedo over Earth's land surfaces (0.0–0.4). *Red* is most reflective, followed by intermediate *yellows* and *greens*, *blues* and *violets* indicate relatively dark surfaces, *white* areas have no data available. No values for the oceanic areas are provided. Epoch of image is 7–22 April 2002. Source: NASA Visible Earth, http://visibleearth.nasa.gov, image credit, Crystal Schaaf, Boston University, based upon data processed by the MODIS Land Science Team

- Atmospheric reflection and scattering are included.
- The orientation and surface properties of the satellite affect the force.

Other factors such as variable cloud cover will also compound the problem. Following Montenbruck and Gill (2001), one can sum the acceleration components from the different Earth elements, where j ($j = 1, \ldots, N$) are individual terms, which correspond to the elements dA_j:

$$\ddot{\vec{r}} = \sum_{j=1}^{N} C_R \left(\nu_j a_j \cos\theta_j^{\mathrm{E}} + \frac{1}{4} \in_j \right) P_{\mathrm{e}} \frac{A}{m} \cos\theta_j^{\mathrm{s}} \frac{dA_j}{\pi r_j^2} \vec{e}_j, \tag{45}$$

where the orientation of the force from element to satellite is determined by $\vec{e}_j$, which is the unit vector pointing from the satellite to the Sun, ν_j is the Earth element shadow function and θ_j^{E} and θ_j^{s} are the surface normal angles of Earth and satellite to the incident radiation, respectively.

9.3.3.8 Other Forces $\left(\vec{f}_{\mathrm{Other}}\right)$

Other non-gravitational forces affect the satellite orbit, such as Yarkovsky thermal drag. The Yarkovsky thermal drag occurs due to hemispheric temperature imbalances between the satellite's Earth facing and opposite hemispheres. Other effects have been investigated as well, some of which are difficult to model properly as they depend to some extent on the spin-orientation of the satellite. More details can be had from Rubincam (1988, 1990) and Scharroo et al. (1991).

9.3.3.9 Empirical Forces $\left(\vec{f}_{\mathrm{Emp}}\right)$

Empirical forces can be modelled to account for small unmodelled forces in the POD software being used. Nine parameters can thus be included in the orbital estimation as solve-for parameters; they take the form (Montenbruck and Gill 2001)

$$\ddot{\vec{r}} = \vec{E} \left(\vec{a}_0 + \vec{a}_1 \sin\nu + \vec{a}_2 \cos\nu \right) \tag{46}$$

and are modelled as once-per-cycle-per-revolution accelerations, where $\vec{a}_0$ is a constant acceleration bias, $\vec{a}_1$ and $\vec{a}_2$ are the coefficients describing the 1 CPR components and the orbit true anomaly is denoted by ν. The empirical accelerations are normally applied to the radial, transverse and normal components (RTN) in a local orbital frame and need to be transformed back into ECI through matrix $\vec{E}$. These empirical accelerations can absorb mismodelling, but can also produce degraded estimates of other parameters, even though the rms error of the O–C residuals seems to be smaller. Typically they are of the order of $\approx 10^{-12}$ m s^{-2}, or less. Larger values will influence other parameters (e.g. gravitational zonal rate value for $J_2 \sim 3 \times 10^{-11}$ year^{-1}).

9.4 Calculated Range

During the processing of SLR data, a calculated range must be found, which uses some orbit propagation routine which includes the previously discussed perturbations, starting with an a priori satellite coordinate and velocity. A variety of integration methods are used in orbit calculation routines, such as Runge–Kutta, multistep and extrapolation methods. These will not be discussed here.

Before the calculated range can be determined, the station position needs to be corrected for certain variations. The station position must be adjusted to its position at the epoch of the SLR measurements, utilising a procedure such as already described in (19). Other adjustments to the station position include adjustments for Earth-tide, ocean loading, atmospheric loading and pole-tide. These all affect the position of the station and produce a three-dimensional displacement of the SLR station, with the effect that even if the ITRF coordinates have been adjusted for plate velocity, the "true" coordinates at the epoch of SLR ranging data being captured have not been found. In addition, a station's position can be affected by atmospheric loading (0.5–3 mm) and mass redistribution in continental water reservoirs, e.g. groundwater, soil moisture and snow.

The majority of these position disturbing processes are described in detail in McCarthy and Petit (2003), in the section on the displacement of reference points (as updated). Updates to the IERS conventions are detailed (containing all corrections to IERS 2003) at http://tai.bipm.org/iers/convupdt/convupdt.html. Different ocean-tide models are available at http://www.oso.chalmers.se/loading.

One can compute the ocean-tide loading displacements for a given SLR site, given the amplitude A_{cj} and phases ϕ_{cj}, where $1 \leq j \leq 11$, as generated by the website. This is easily implemented utilising FORTRAN code (hardisp.f) (developed by Duncan Agnew, University of California, San Diego, La Jolla), which determines local dU, dS, dW displacements considering a total of 342 constituent tides using a spline interpolation of the tidal admittances. The model precision is about 0.1%. Amplitudes for ocean loading vary from millimetre to the few centimetre level. The HartRAO SDAS utilises hardisp.f, but we have converted it from FORTRAN to C.

Solid Earth-tides can cause a displacement of tens of centimetres at certain locations, and displacement due to pole-tide can amount to several centimetres and has to be accounted for. A complete description is given in the IERS conventions for both Earth-tide and pole-tide.

Once these adjustments to the SLR station position have been made, the range is found through iterative solutions of two light-time equations for the uplink and downlink paths, and this is fully described in Montenbruck and Gill (2001). For the uplink path a fixed-point iteration with

$$\tau_u^{(i+1)} = 1/c \cdot \left| \vec{r}\,(t - \tau_d) - \vec{R}\left(t - \tau_d - \tau_u^{(i)}\right) \right| \tag{47}$$

is executed in a loop until τ_u reaches a predefined accuracy threshold. Similarly for the downlink, starting from an initial value of $\tau_0 = 0$, consecutive determinations are done using the fixed-point iteration

$$\tau_d^{(i+1)} = 1/c \cdot \left| \vec{r}\left(t - \tau^i\right) - \vec{R}\left(t\right) \right| . \tag{48}$$

The two-way range is then determined from the average of ρ_u and ρ_d, so that

$$\rho = 0.5\left(\rho_u + \rho_d\right) . \tag{49}$$

In (47) and (48), $\overrightarrow{r}$ is the satellite position vector and $\overrightarrow{R}$ is the SLR station position vector.

Finally then, after calculating the range, the observed–computed (O–C) residuals can be formed in a least squares solution where the force model parameters are adjusted against the observation model as determined by the SLR range observations. An example is given in Fig. 9.8, a screen capture of the HartRAO SDAS. Manipulation of the number of solve-for parameters, a priori set-up values and processing strategy will change these values as solve-for of orbital, geophysical and empirical parameters are adjusted. Additional examples from the ILRS processing centres can be found at the ILRS website.

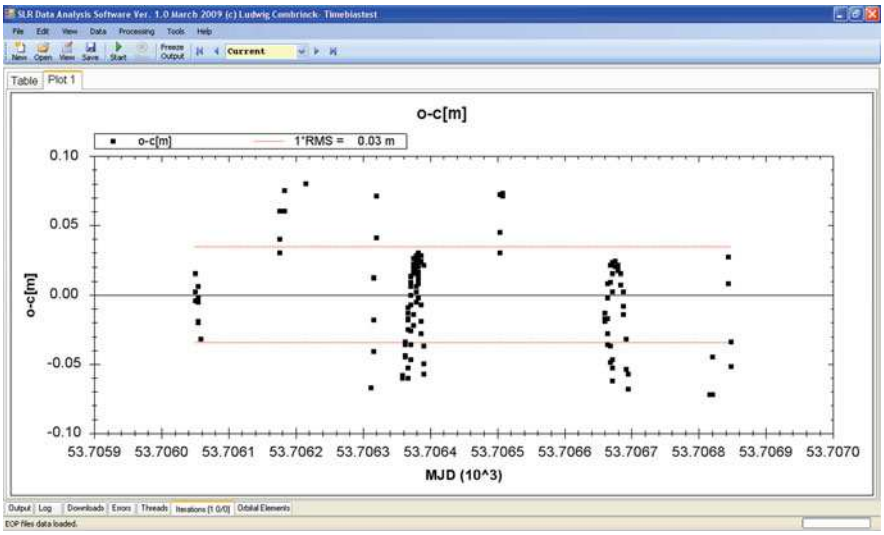


Fig. 9.8 Example of observed–computed residuals; a 1-day arc of LAGEOS 2, clearly showing unmodelled effects, where the range as observed and the range as calculated disagree; rms value of fit is ~3 cm

9.5 SLR System and Logistics

There are many different system configurations and each SLR station will have instrumental, configuration and system-specific set-ups which will differ in detail from other stations. The exception, perhaps, to some degree is contained within the eight-station NASA SLR network that includes five Mobile Laser Ranging Systems (MOBLAS 4–8). Other NASA network SLR systems are the two Transportable Laser Ranging Systems (TLRS 3 and 4) and the McDonald Laser Ranging System (MLRS), which are also used for Lunar Laser Ranging (LLR).

The MOBLAS stations are located at Monument Peak, California (MOBLAS-4), Yarragadee, Australia (MOBLAS-5), Hartebeesthoek, South Africa (MOBLAS-6), Greenbelt, Maryland, at the Goddard Space Flight Center (MOBLAS-7 as well as TLRS-4), and Tahiti, French Polynesia (MOBLAS-8). Other NASA stations are in Arequipa, Peru (TLRS-3), and the LLR capable station near Ft. Davis, Texas (MLRS). These stations operate in a global network, the International Laser Ranging Service (ILRS), which has certain station and data quality requirements before allowing a station to become a member of the network. In Sect. 9.5.1 a quick review is made of a typical SLR system to describe some of the main subsystems. Thereafter (Sects. 9.6.1 and 9.6.2), some network and logistical issues are discussed very briefly.

9.5.1 System Configuration

The system configuration, software and operational procedures for the MOBLAS systems are the same to a large degree. Even so, there are times when some MOBLAS stations have received upgrades and others are still in pre-upgrade configuration. It is probably fair to say that all SLR stations are unique in one way or another. These sub-assemblies as depicted in Fig. 9.9 are typical of SLR systems, and most will be found in all SLR systems in one form or another.

Typical assemblies are

- *Laser assembly*: the laser table and peripheral equipment, such as cooling system, power supplies and monitoring equipment
- *Tracking and mount control*: telescope optics and telescope pointing and steering
- *Data measurement*: measurement and Laser Range Controller (LRC) logic
- *Timing*: produces timing and synchronisation signals
- *Controller*: PC controlling all the subsystems
- *Processor*: PC for processing data and sending or receiving files
- *Safety*: safety features integrated into the system

9.5.1.1 Laser Assembly

Different laser assemblies are found throughout the ILRS network and limited space precludes discussing all of them, so MOBLAS-6 is used as an example. The laser

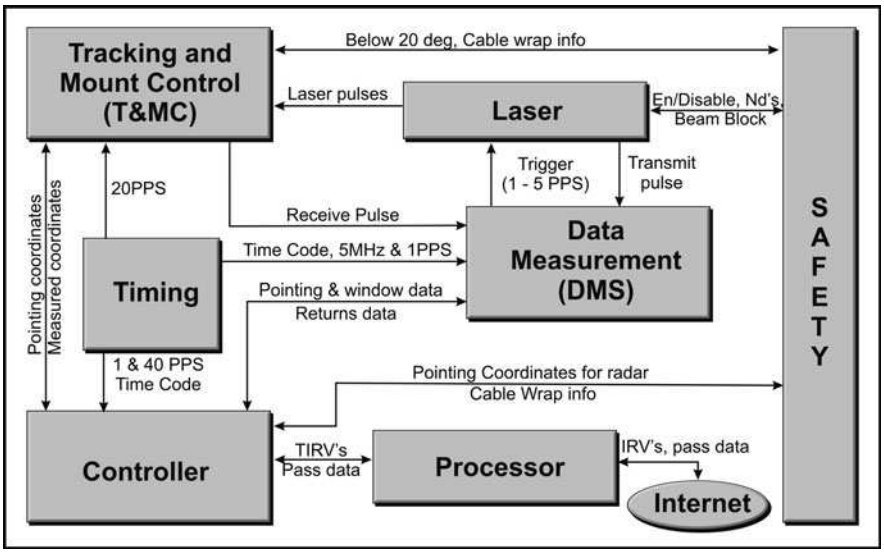


Fig. 9.9 Basic sub-assemblies of MOBLAS-6

is located inside the mobile trailer, but isolated from the structure and located on a massive pier tied to bedrock. The isolation is of such a nature that any movement of the pad on which the trailer is located, or any motion of the trailer itself, does not adversely affect the stability (and position) of the laser table. In Fig. 9.10 the current optical table layout of the MOBLAS-6 is depicted. This system is cavity dumped and has active/passive mode locking. The laser head is component [6a], which consists of a liquid cooled Nd:YAG rod and a linear flash-tube assembly, which pumps the rod. Pumping generates a family of pulses, and one of these pulses needs to be selected for further amplification and processing through the laser system. A pulse-slicer [11a] is triggered by an adjustable detector [11b], which provides a signal for the avalanche switch [11d]. The avalanche switch in turn excites a crystal in the pulse-slicer [11a], causing a phase shift suitable for reflection at [15c] (polariser).

Thereafter the polariser outputs the phase-shifted pulse from the cavity resonator made up of components between [18a] and [8]. Mode locking is accomplished by a radio frequency (∼29 MHz) modulated crystal [7]. Calibration is done via He–Ne laser [1], which is routed through the optics via mirrors [2a] and [2b], allowing accurate optical alignment. After reflection at the polariser, the beam is directed through a beam expander [10], where the beam is expanded slightly to match the larger diameter of the rod in the amplifier assembly [17].

The maximum output of this system is 120 mJ at 532 nm, at a repetition rate of 5 Hz. A zero-degree turning mirror [18b] provides for double-pass (pumped twice) amplification. Consequently the polariser [15d] selects a pulse to be passed on to the frequency doubler [20], which converts the infra-red (1,064 nm) to green light

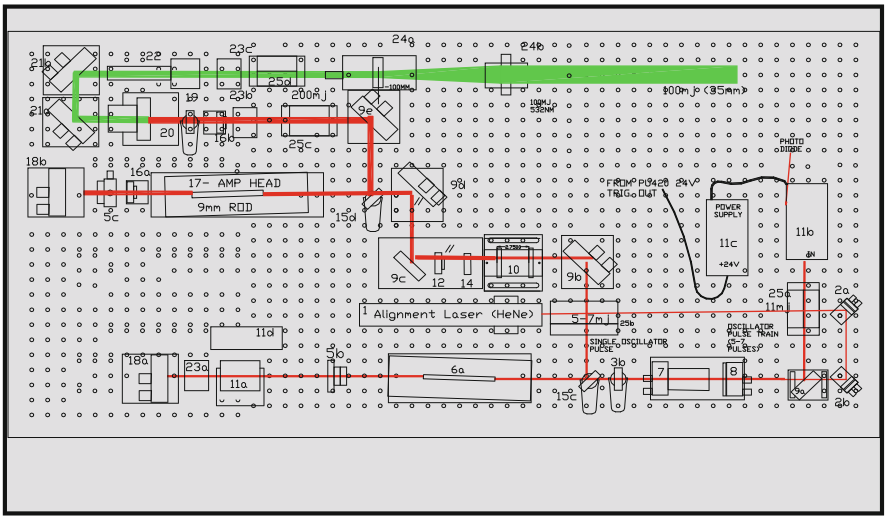


Fig. 9.10 Schematic of the Nd:YAG laser of MOBLAS-6, illustrating the main components. Current power output of this system is 120 mJ at a pulse rate of 5 Hz

(532 nm). Efficiency of the frequency doubler is about 50%, so that a large amount of energy is lost at this point in the optical train.

Two dichroic mirrors (a and b) reflect the green light mostly; this light is then detected by a photodiode [22] which provides the electrical pulse for the start-of-flight measurement (to determine the transmit epoch a phase delay has to be added). Consequently the laser pulse is passed through a beam expander, which expands the laser beam to approximately 100 mm, from where it passes via two mirrors into the Coudé path. This Coudé path allows the laser pulse to be routed to the exit optical assembly (another beam expander) via an assembly of mirrors (some which rotate with the elevation and azimuth axes), which is mounted (in the case of MOBLAS-6) on the side of the main tube of the 0.76-m telescope.

9.5.1.2 Tracking and Mount Control

Similar to many other SLR systems, such as the Borowiec SLR located in Poznan, Poland, MOBLAS-6 (Fig. 9.11) has an azimuth–elevation type mount with separate transmit and receive optics. The receive optics consists of a 76.2-cm main mirror in Cassegrain configuration and a photomultiplier tube located at the focal point of the hyperbolic subreflector (secondary mirror). The primary reflector is an *f*/1.5 parabolic mirror. In total (primary plus secondary) the telescope is an *f*/5 system. Also attached to the main telescope is a guide telescope; this is a 127-mm (*f*/10) Schmidt–Cassegrain reflector. The guide telescope is equipped with an illuminated reticle and a CCD camera; it can therefore be used remotely by the SLR operator to position the receive telescope for tracking or collimation. Similar arrangements have been used to create models for mount errors (Schillak 2004).

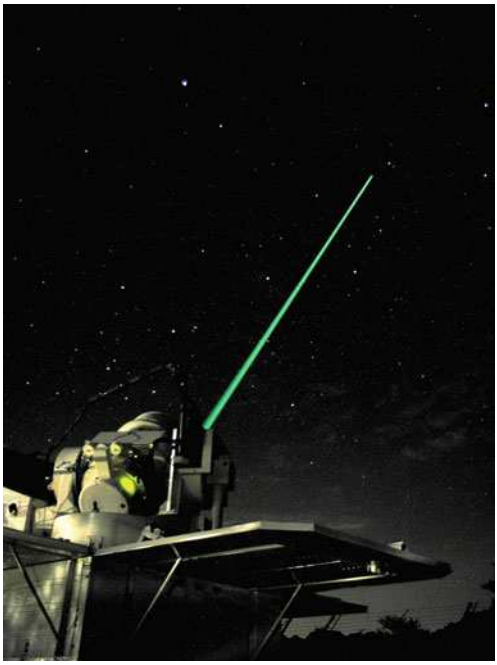


Fig. 9.11 Moblas-6 ranging to a satellite during night time; the 532-nm laser is clearly visible due to the camera integrating the individual 200-ps pulses. Photo credit: Michael Gaylard, Hartebeesthoek Radio Astronomy Observatory

Consolidated Prediction Format (CPF) (previously Inter-range Vectors, IRVs) satellite position predictions are provided to enable an SLR system to track a given satellite. These data are converted and used to provide azimuth and elevation positioning signals which are in turn applied to the T&MC subsystem servo interfaces. The Mount Positioning And Control System (MPACS) readout chassis in the T&MC subsystem is a signal conditioning system that processes the inductosyn inputs. This readout chassis generates corresponding azimuth and elevation angle information for display. Another unit, the Servo Control Chassis, receives inputs from the controller computer and inductosyn assemblies, whereupon it generates the corresponding drive signals to position the mount. Positioning is done by two independent drive motors on azimuth and elevation axes, using direct-drive, brushless d.c. motors via class A amplifiers.

9.5.1.3 Data Measurement

The Data Measurement Subsystem (DMS) is synchronised with the station's standard frequency generator (timing subsystem in Fig. 9.12). Detection and recording of the time interval between the transmit pulse and the reflected receive pulse are the main functions of the DMS. In addition it analyses transmit and receive pulses and provides the required outputs to the controller computer for ranging computations and recording. The DMS interfaces with other subsystems to provide the controller subsystem computer with the data required to resolve the range measurement. The Laser Ranging Control (LRC) is contained within the DMS and provides the electrical pulses to the laser subsystem which enables a 5-Hz firing rate.

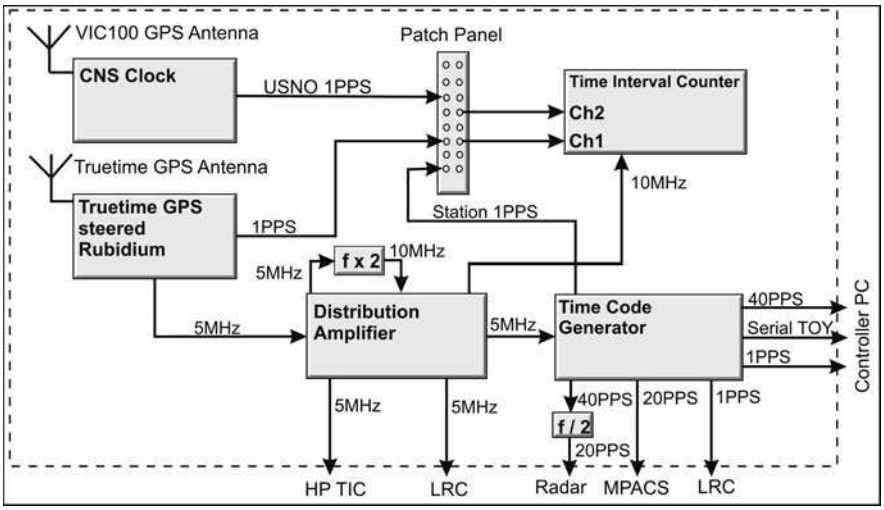


Fig. 9.12 The timing subsystem of MOBLAS-6. A 5-MHz signal is distributed to various subsystems which require accurate time and frequency

9.5.1.4 Timing

Time and frequency are critical components of any SLR system. At HartRAO we have access to a Hydrogen MASER clock used for VLBI, but for SLR purposes a rubidium frequency standard (typically excellent short-term stability ($\sim 2 \times 10^{-11}$ Allan variance (1 s)) and long-term drift ($\sim 5 \times 10^{-11}$/month)) steered by GPS are utilised.

In order to count the time interval between receive and transmit, a time interval counter (TIC) is used; some stations utilise the Stanford SR620 TIC units, the MOBLAS stations use a HP5370B and many other TIC units have been and are in use by other stations. Figure 9.12 describes the timing subsystem. A Truetime rubidium frequency standard is maintained close to UTC as its long-term drift is corrected regularly via GPS; this frequency standard provides a 1-pulse-per-second (PPS) and 5-MHz signal. The 1 PPS signal is used to compare the SLR station time to the CNS clock (also steered by GPS) and to determine the system delay through the distribution amplifier and time code generator via the Agilent 53131 TIC. The SLR system utilises the 5-MHz reference signal via the distribution amplifier which maintains an adequate signal level throughout the system. A time code generator utilises the 5 MHz to produce time code data in a standard NASA 35-bit 1-s format as well as 1, 20 and 40 PPS. Various subsystems are fed from the timing subsystem, including the radar (slaved with the SLR telescope) and the Laser Ranging Control circuits. Very precise event timers have been developed recently (which time tags the transmit and receive epochs instead of counting the time interval), e.g. Samain et al. (2007) report the development of an event timer with a precision better than 2 ps, linearity in the range of 1 ps and thermal drift of below 1 ps/ °C.

9.5.1.5 Controller

The controller subsystem is a real-time, computer-controlled data handling system that controls the T&MC subsystem servo control section to position the SLR telescope mount for ranging operations. In addition, the controller computes, formats and records the ranging data based on the computer program and inputs from the other subsystems. Mount positional data from the shaft angle encoders are compared with the programmed positioning data. The controller subsystem makes the comparison and generates a correction (error) signal that it sends to the Servo Interface Chassis to reposition the mount.

9.5.1.6 Processor

On-site data analysis is performed by the processor subsystem. Data received (required for accessing the satellites) from NASA SLR headquarters are processed, and SLR data are returned for upload to the Crustal Dynamics Data Information System (CDDIS) (Noll and Dube 2004) and other data banks where these data are accessible for scientific community use. After a pass is complete, the controller computer transmits the pass data to the processor. The processor then converts the data into appropriate formats and performs post-pass analyses. The performance analyst can use this information to determine the quality of the pass data, how well the equipment operated and how to correct malfunctions.

9.5.1.7 Safety

Some SLR systems are eye-safe, such as the newly developed NASA SLR2000 system (Degnan 1994). However, most current systems are not eye-safe due to the high-power lasers used for ranging. Based on ANSI Z136.1 – 1986 requirements, the MOBLAS laser beams are capable of causing blind spots in unprotected eyes at up to 39 km. Safety is therefore extremely important and the system has been designed to ensure eye safety. Two main safety components are integral to the system: the safety interlock and its peripheral equipment and a radar system slaved with the SLR telescope. The use of motion sensors, pressure pads, mount position sensors and a video camera ensures operator and other personnel safety. Aircraft are detected by the radar system and disable the laser upon detection.

9.6 Network and International Collaboration

Satellite laser ranging is an international effort and is such a large undertaking that a single country cannot operate and maintain a global network; therefore international cooperation and collaboration is of paramount importance. Currently there are 22 operational and 16 associate stations in the ILRS network. There are more than these 38 SLR stations; however, all of them are not part of the ILRS network. In order to

qualify as an operational station, a set of criteria relating to number of passes and tracking accuracy need to be met over a specified period.

9.6.1 Tracking Network

The distribution of the SLR network is not uniform, which results in geometrical and coverage weaknesses, with more stations located in the northern hemisphere than the southern hemisphere. In addition, data quantity is station dependent, with some stations routinely providing a large volume of high-quality data, while others may produce less data based on weather conditions, staffing levels and operational status. A map of the current network is presented in Fig. 9.13; this map is available on the ILRS website (http://ilrs.gsfc.nasa.gov) and is maintained by NASA personnel (Noll and Torrence 2008).

9.6.2 International Laser Ranging Service

The ILRS (Pearlman et al. 2002) is one of the space geodetic services of the International Association of Geodesy (IAG) and "provides global satellite and lunar

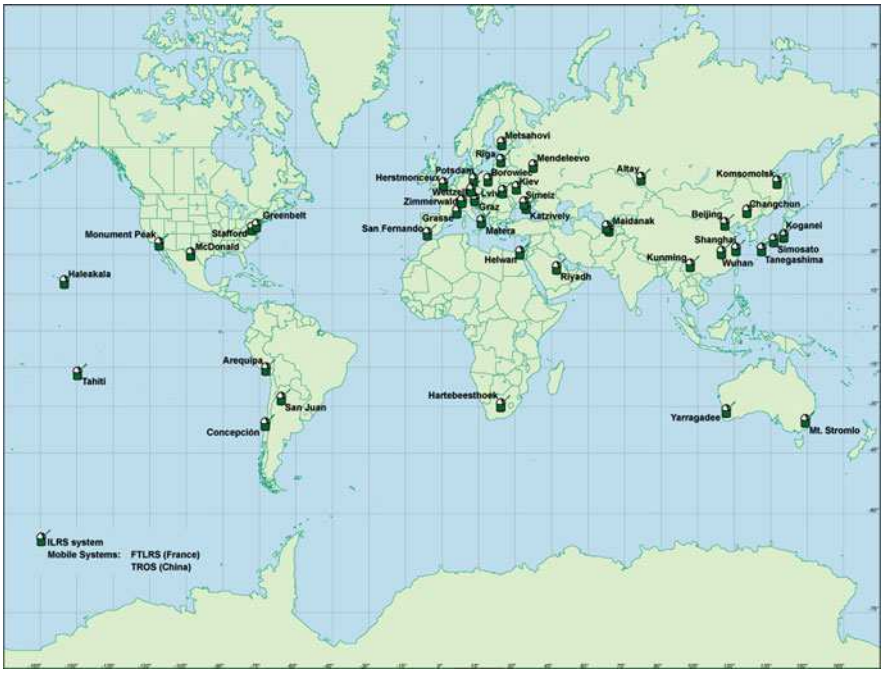


Fig. 9.13 Map of the distribution of the ILRS SLR network (Source: http://ilrs.gsfc.nasa.gov)

laser ranging data and their related products to support geodetic and geophysical research activities as well as IERS products important to the maintenance of an accurate International Terrestrial Reference Frame (ITRF)". A comprehensive and very well maintained website can be found at http://ilrs.gsfc.nasa.gov. This IAG service is the main hub of the different activities of the laser ranging community.

9.7 Summary

Satellite laser ranging makes a unique contribution to space geodesy by being the only optical space geodesy technique with 1- to 2-cm ranging accuracy capability. Uses of the technique are expanding beyond the regular tracking of Earth satellites or the tracking of the Moon, through the development of laser transponders, which will soon allow regular tracking and other applications (e.g. optical communications and time transfer) in the solar system. Products such as Earth orientation parameters, station position and velocity, as well as orbit calibration facilitate a large range of applications in diverse fields of orbital dynamics, geophysics and geodynamics, reference frame establishment and fundamental physics. Utilising range data determined from the time-of-flight of a laser pulse from an SLR station to a satellite and back, it is possible to adjust force model parameters and estimate a range of parameters which are of interest to different fields of science and applications. In combination with the other space geodetic techniques, SLR is contributing to the maintenance and improvement of the ITRF. The SLR technique continues to be improved in terms of observational and technical aspects, working towards a sub-centimetre accuracy level. Active working groups continuously evaluate and improve the analysis and interpretation of its observables. Furthermore, it is an integral component of the Global Geodetic Observing System (GGOS), a project of the International Association of Geodesy (IAG) through its contributions to the three fundamental geodetic observables and their variations – the shape of the Earth, its gravity field and rotational motion. The ILRS plays an instrumental and guiding role through setting tracking priorities and standards; it also provides channels of communication and promotes international collaboration through working group activities and international meetings.

Acknowledgments The data used in this work were acquired as part of NASA's Earth Science Data Systems and archived and distributed by the Crustal Dynamics Data Information System (CDDIS). Acknowledgement is made of the use of NASA MOBLAS system description documentation for Figs. 9.9, 9.10 and 9.12.

References

Anderson, P., Lehr, C., Maestre, L. and Snyder, G. (1966) 2A10-Laser experiments for determining satellite orbits. IEEE J. Quantum Electron., 2(8), 215–219

Arnold, D.A. (1978) Optical and infrared transfer function of the LAGEOS retroreflector array. Final Report, NASA Grant NGR 09-015-002, Smithsonian Institution Astrophysical Observatory, Cambridge

Artyukh, Y. (2007) Riga event timers: principles of operation and performance characteristics. Presentation at ILRS Fall 2007 Workshop, 25–28 September 2007, Grasse

Bonnefond, P., Exertier, P., Schaeffer, P., Bruinsma, S. and Barlier, F. (1995) Satellite altimetry from a short-arc orbit technique: application to the Mediterranean. J. Geophys. Res., 100(C12), 25365–25382

Ciufolini, I. and Wheeler, J.A. (1995) Gravitation and Inertia. Princeton University Press, Princeton

Combrinck, L. and Suberlak, V. (2007) Earth-tide as parameter of crustal motion correction for SLR station displacement. South Afr. J. Geol., 110(2–3), 203–210

Degnan, J.J. (1993) Millimeter accuracy satellite laser ranging: a review. In: Smith, D.E. and Turcotte, D.L. (eds) Contributions of Space Geodesy to Geodynamics: Geodynamics Series, Vol. 25. AGU, Washington, pp. 133–162

Degnan, J., (1994) SLR2000: an autonomous and eyesafe satellite laser ranging station. Proceedings of the 9th International Conference on Laser Ranging Instrumentation, Canberra, November 7–11, pp. 312–323

Dunn, P., Torrence, M., Kolenkiewicz, R. and Smith, D. (1999) Earth scale defined by modern satellite ranging observations. Geophys. Res. Lett., 26(10), 1489–1492

Eanes, R.J., Schutz, B. and Tapley, B. (1983) Earth and ocean tide effects on Lageos and Starlette. In: Kuo, J.T. (ed) Proceedings of the 9th International Symposium on Earth Tides. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, pp. 239–250

Exertier, P., Nicolas, J. and Barlier, F. (2000) SLR: a point of view on scientific achievements and future requirements. Presentation at the 12th International Workshop Laser Ranging, Matera, November 2000

Henize, K.G. (1957) The Baker-Nunn satellite tracking camera. Sky Telescope, 16(10), 108–111

Hoffman-Wellenhof, B. and Moritz, H. (2005) Physical Geodesy. Springer, Wien

Huang, C. and Liu, L. (1992) Analytical solutions to the four post-Newtonian effects in a near-Earth satellite orbit. Celest. Mech. Dynam. Astron., 53, 293–307

Husson, V.S. (1993) Historical MOBLAS System Characterisation. Located at http://ilrs.gsfc.nasa.gov

Kaplan, G.H. (2005) The IAU Resolutions on Astronomical Reference Systems, Time Scales and Earth Orientation Models, Explanation and Implementation. United States Naval Observatory Circular no. 179, USNO, Washington

Klein, B.J. and Degnan, J.J. (1974) Optical antenna gain, 1. Transmitting antennas. Appl. Opt., 13, 2134–2140

Luthcke, S.B., Zelensky, N.P., Rowlands, D.D., Lemoine, F.G. and Williams, T.A. (2003) The 1-centimeter orbit: Jason-1 precision orbit determination using GPS, SLR, DORIS, and altimeter data. Mar. Geod., 26, 399–421

Marini, J.W. (1972) Correction of satellite tracking data for an arbitrary tropospheric profile. Radio Sci., 7, 223–231

Marini, J.W. and Murray, C.W. (1973) Correction of laser range tracking data for atmospheric refraction at elevations above 10 degrees, NASA Technical Memorandum, NASA-TM-X-70555, 60pp

McCarthy, D.D. and Petit, G. (2003) IERS Conventions (2003) (IERS Technical Note; 32). Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, 72–84

Mendes, V.B. and Pavlis, E.C. (2004) High-accuracy Zenith delay prediction at optical wavelengths. Geophys. Res. Lett., 31, L14602

Mendes, V.B., Prates, G., Pavlis, E.C., Pavlis, D.E. and Langley, R.B. (2002) Improved mapping functions for atmospheric refraction correction in SLR. Geophys. Res. Lett., 29, 1414

Montenbruck, O. and Gill, E. (2001) Satellite orbits: models, methods and applications. Springer, Berlin

Noll, C.E. and Dube, M. (2004) Archiving space geodesy data for 20+ years at the CDDIS. Eos Trans. Am. Geophys. Union, 85(47), Fall Meet. Suppl., Abstract SF21A-0010

Noll, C.E. and Torrence, M. (2008) ILRS web site update. Poster Presented at the 16th International Laser Ranging Workshop, Poznan, 13–17 October 2008

Otsubo, T. and Appleby, G. (2003) System-dependent centre-of-mass correction for spherical geodetic satellites. J. Geophys. Res., 108(B4), 2201

Pavlis, D.E., Moore, D., Luo, S., McCarthy, J.J. and Luthcke, S.B. (1999) GEODYN Operations Manual: 5 Volumes. Raytheon ITSS, Greenbelt, December, 1999

Pearlman, M.R., Degnan, J.J. and Bosworth, J.M. (2002) The international laser ranging service. Adv. Space Res., 30, 135–143

Ries, J.C., Huang, C. and Watkins, M.M. (1988) The effect of general relativity on near-Earth satellites in the solar system barycentric and geocentric reference frames. Phys. Rev. Lett., 61, 903–906

Rubincam, D.P. (1988) Yarkovsky thermal drag on Lageos. J. Geophys. Res., 93(B11), 13805–13810

Rubincam, D.P. (1990) Drag on the LAGEOS satellite. J. Geophys. Res., 95, 4881–4886

Rubincam, D.P. and Weiss, N.R. (1986) Earth albedo and the orbit of LAGEOS. Celestial Mech., 38, 233–296

Samain, E., Torre, J.M., Guillemot, P., Leon, S., Petitbon, I. and Vrancken, P. (2007) OCA event timer. Presentation at ILRS Fall 2007 Workshop, 25–28 September 2007, Grasse

Sanchez, B.V. (1974) Rotational dynamics of mathematical models of the nonrigid Earth. Applied Mechanical Research Laboratory Report. 1066, The University of Texas at Austin, Austin

Scharroo, R., Wakker, K.F., Ambrosius, A.C. and Noomen, R. (1991) On the along-track accelerations of the Lageos satellite. J. Geophys. Res., 96, 729–740

Schillak, S. (2004) Analysis of the process of the determination of station coordinates by the satellite laser ranging based on results of the Borowiec SLR station in 1993.5–2000.5. Artif. Satellites, 39(3), 223–263

Sinclair, A.T. and Appleby, G.M. (1986) SATAN—Programs for the determination and analysis of satellite orbits for SLR data, SLR Technical Note, No. 9, Royal Greenwich Observatory, Herstmonceux, 14pp

Smith, D.E. and Dunn, P.J. (1980) Long term evolution of the LAGEOS orbit. J. Geophys. Res. Lett, 7, 437–440

Smith D, Zuber, M., Torrence, M., McGarry, J. and Pearlman, M. (2006) Laser ranging to the Lunar Reconnaissance Orbiter (LRO). Presentation at the 15th International Workshop on Laser Ranging Instrumentation, 15–20 October 2006, Canberra

Smith, M.A. (1964) Laser Tracking Success. Published in Flight International, Chief editor M.A. Smith, Number 2908, Vol. 86

Solomon, L.H. (1967) Some results at Baker-Nunn tracking stations. Smithsonian Astrophysical Observatory Special Reports, No. 244, 14pp

Standish, E.M. (1998) JPL Planetary and Lunar Ephemerides, DE405/LE405, JPL IOM 312.F-98-048

Tapley, B.D., Schutz, B.E., Eanes, R.J., Ries, J.C. and Watkins, M.M. (1993) LAGEOS laser ranging contributions to geodynamics, geodesy, and orbital dynamics, in contributions of space geodesy to geodynamics: Earth dynamics. Geodyn. Ser., 24, 147–174

Tapley, B.D., Schutz, B.E. and Born, G.H. (2004) Statistical Orbit Determination Burlington. Elsevier Academic Press, San Diego

Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F., Gunter, B., Kang, Z., Nagel, P., Pastor, R., Pekker, T., Poole, S. and Wang, F. (2005) GGM02 – an improved Earth gravity field model from GRACE. J. Geod., doi: 10.1007/s00190-005-0480-z

Vallado, D.A. (2001) Fundamentals of Astrodynamics and Applications, 2nd Ed. Space Technology Library, Microcosm Press, El Segundo; Kluwer Academic Publishers, Dordrecht

Vrancken, P., Samain, E. and Guillemot, P. (2008) Design and test of the time transfer by laser link (T2L2) optical subsystem. Proc. SPIE, 7003, 700311, doi: 10.1117/12.781123

Chapter 10
Superconducting Gravimetry

Jürgen Neumeyer

Contents

J. Neumeyer (✉)
GeoForschungZentrum Potsdam, Potsdam, Germany
e-mail: Juergen_Neumy@yahoo.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_10,

10.1 Introduction

The objective of gravimetry (Torge 1989) is the determination of the gravity field of the Earth and the surface gravity effects as functions of position and time by measurements of the gravity intensity which is the magnitude of the gravity acceleration g on the Earth's surface. The gravity g and its variation depend on the mass attraction (law of gravitation) according to the arrangement of the terrestrial and exterritorial masses (celestial bodies) and the Earth rotation (centrifugal acceleration). Mass distribution and rotation are subject to variations in time. Examples for gravity changes δg are

- Differences due to the position (equator–pole): $\sim 5 \times 10^{-2}\mathrm{m/s^2}$
- Differences due to the elevation (high mountains–deep sea): up to $\sim 5 \times 10^{-2}\mathrm{m/s^2}$
- Periodic tidal effects: up to $\sim 3 \times 10^{-6}\mathrm{m/s^2}$
- Mass redistribution in the atmosphere: up to $\sim 2 \times 10^{-7}\mathrm{m/s^2}$
- Long-term terrestrial mass displacements: in the order of $\sim 10^{-7}\mathrm{m/s^2}$

Two principal measurement methods are in use: absolute (g) and relative (δg) gravity measurements. For absolute gravity measurements the free fall method is mostly applied, e.g. the FG5 absolute gravimeter (Niebauer et al. 1995). The mass-spring system is the most common one for relative gravity measurements, e.g. LaCoste and Romberg (Torge 1989) or Scintrex (Timmen and Gitlein 2004) spring gravimeters, and the superconducting gravimeter (SG) where the spring is replaced by a magnetic suspension of the proof mass having the form of a hollow sphere.

The superconducting gravimeter is presently the most sensitive relative instrument with the lowest drift rate which can be modelled and corrected. Therefore, it is used for measuring and detecting weak gravity effects in a period range from minutes to years.

The SG is, like mass-spring gravimeters, a gravity sensor measuring gravity variations, δg, associated with mass movements in its near and far surroundings. According to the law of gravitation the gravity sensor, a proof mass, is sensitive to

changes of and squared distance changes to the source mass. Because matter has the properties of gravity and inertia, the gravity sensor (a proof mass) reacts to time variations of

1. Gravitational forces (Newtonian attraction) caused by redistribution and density variations of all surrounding masses
2. Inertial forces caused by accelerations, i.e. the second time derivative of the vertical position of the gravimeter site

Therefore, the SG recordings contain not only the tidal gravity from the Moon and Sun as well as from other celestial bodies, but also the gravity effects induced by various geophysical and geodynamic sources on global, regional and local scales.

The gravimeter measures the sum of gravity changes caused by these different sources. It cannot separate them. For separating the different gravity effects special models, analysis methods and additional data are necessary, e.g. meteorological and hydrological data. In many cases research is focused on global gravity effects such as Earth tides, seismic normal modes, core modes, nearly diurnal-free wobble (NDFW) and Chandler wobble. To investigate one individual effect, all the other components have to be removed from the gravity data. The other constituents are disturbing signals in this case.

One disturbing contribution consists of accelerations (vibrations) usually considered as noise (seismic, industrial and ocean noise), which can be reduced by low- or band-pass filtering (if the particular frequency bands are known). Supplementary instrumental effects (drift, offsets and instrumental noise) superimpose the gravimeter signal. Most of these effects can also be removed from the data. Another contribution usually treated as disturbing signals are environmental influences. To remove them, they have to be modelled and hence accurate data of the atmosphere and hydrosphere (e.g. atmospheric pressure, groundwater level, soil moisture) must be available. For investigation of one individual gravity effect the other gravity effects must be removed from the gravity data precisely. The better the different gravity effects can be modelled and reduced from the gravity data, the better the remaining individual effect can be investigated because it is less or not superposed by the other effects.

In Table 10.1 the main global surface gravity effects recorded by the SG are summarized. All these effects are included in the raw gravity data. Depending on the SG location, additional local or regional gravity effects can be contained in the gravity signal mainly caused by the hydrosphere, or other effects like co- and post-seismic gravity changes, crustal deformation in tectonic active zones, postglacial rebounds.

Besides measuring, analysing and interpreting the above surface gravity effects, SG gravity measurements and analysis results can be used for

- Earth tide reduction for relative and absolute gravity measurements and other precise measurements like satellite positioning, GPS, laser ranging or radiointerferometric methods
- Verification of global Earth tide models

- Validation of satellite-derived gravity variations
- Validation of global and regional ocean tide models
- Validation of global and regional hydrological models
- Investigation of changes in the tidal parameters

Further applications for SG measurements are

- Pre-, co- and post-seismic gravity changes
- Studying the local, regional and global hydrological cycle in combination with, e.g., satellite-derived gravity variations and hydrological models
- Gravity changes in tectonic active areas
- Silent earthquakes
- Volcanic activities
- Post-glacial rebounds etc.

Table 10.1 Main surface gravity effects (period ranges and gravity effects are average peak indicators depending on site and event)

Period range	Physical source	Gravity effect
~0.1–10 s	Microseismic (natural or man-made) noise	Up to 10 μGal (10^{-8} m/s^2)
~0.1–100 s	Earthquakes	Up to multiple measurement range of SG
Minutes to 1 h	Earth's free oscillation	<1 μGal
4–8 h (expected)	Translational oscillation of the inner core (Slichter modes)	<0.1 nGal (10^{-11} m/s^2) expected
6 h–1 year	Body tides	Up to 280 μGal
6 h–1 year	Tidal ocean loading	Up to 10 μGal
Hours to years	Non-tidal ocean loading	Up to 2 μGal
Minutes to years	Atmospheric pressure variations	Up to 20 μGal
Minutes to years	Groundwater level and soil moisture variations	Up to 10 μGal/m
Up to ~435 days	Polar motion	Up to 10 μGal
~430 days	Earth's nearly diurnal-free wobble (NDFW)	Affects Earth tides
Months to years	Different sources of secular gravity variations	μGal range
Months to years	Sea level changes	<0.1 μGal range

Measuring and analysing of these effects are now directly addressed to the "Global Geodynamic Project" (GGP) (Crossley et al. 1999; Crossley 2004; Hinderer and Crossley 2004), a network of presently 23 worldwide distributed SG stations (Fig. 10.1) equipped with similar hardware and software for data acquisition. It is in operation since July 1997. More information about the status

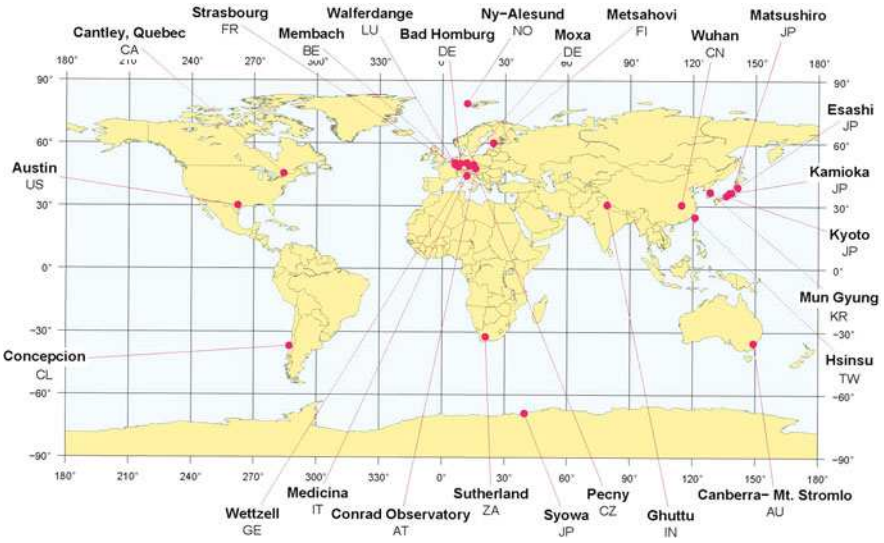


Fig. 10.1 GGP SG network

of the SG stations and the instruments can be found in the GGP home page (http://www.eas.slu.edu/GGP/ggphome.html). The recorded SG and environmental data are available in the GGP database (http://ggp.gfz-potsdam.de).

10.2 Description of the Instrument

The first fundamental design of a superconducting gravimeter was reported by Prothero and Goodkind in 1968. With this operational prototype the SG has been deployed for application in science. This design principle is still in use. However, numerous modifications and improvements have been made by GWR Instruments, San Diego, California, to achieve the present performance of the SG (Warburton and Brinton 1995; Richter and Warburton 1989; Warburton et al. 2000; Hinderer et al. 2007). Figure 10.2 shows the components of the SG: the liquid helium-filled dewar with the gravity sensing unit (GSU) (1) and the coldhead (2), the flexible tubes to the helium liquefying compressor (3) and the data acquisition and control electronics (4).

The superconducting gravimeters belong to the family of relative mass-spring gravimeters. In contrast to the classical mass-spring gravimeter, e.g. LaCoste and Romberg (LCR) (Torge 1989), the SG has no spring. It is replaced by a "virtual spring design" using magnetic levitation of the proof mass rather than a spring. Therefore, the SG performance is not limited by the disadvantages inherent to mechanical springs and it almost completely solves the drift problem having mechanical mass-spring systems (SG some microgalileos per year (nearly linear), e.g. LCR several microgalileos per month).

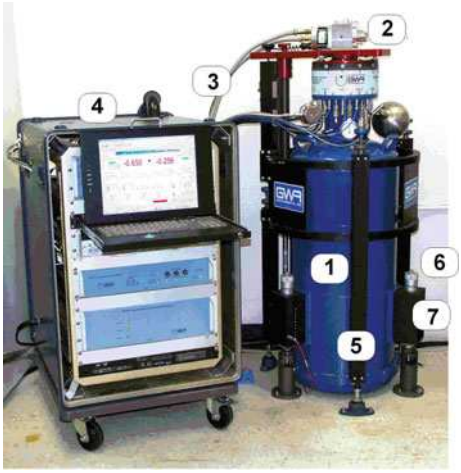


Fig. 10.2 Components of the SG (photo from GWR Instruments). 1 = dewar with gravity sensing unit (GSU); 2 = coldhead SHI SDRK 101; 3 = flexible tubes to the compressor; 4 = electronic rack with data acquisition and control electronics; 5 = coldhead support frame; 6 = micrometer; 7 = thermo-mechanical levellers

10.2.1 Gravity Sensing Unit

The basic element of the gravity senor of the SG is a superconducting sphere housed in the gravity sensing unit (GSU) (Fig. 10.3), which is located in a helium bath of a super-insulated dewar (Fig. 10.2). The superconducting sphere made from niobium (diameter = 2.54 cm, weight = about 5 g) levitates in an ultra-stable magnetic field generated by the *lower and the upper superconducting coils* made from niobium wire. The current in the coils is permanently constant and the resistance of the coils is practically zero as long the superconducting coils remain below their superconducting temperature (9.2 K for niobium). On the other hand, the levitation force varies with temperature roughly by 10 μGal/mK (Goodkind 1999). Therefore, the elements of the GSU are in a vacuum can surrounded by the liquid helium bath contained in the SG's dewar. The temperature within the GSU is adjusted and controlled to 4.2 K within a few microkelvin using an electronic control unit including a *germanium thermometer* and the *heater for temperature control*. This regulated temperature is about 0.1 K above the liquid helium bath. Because of these operating conditions the SG gravity sensor is almost completely isolated from environmental effects caused by changes in external temperature, humidity and atmospheric pressure in contrast to the mechanical mass-spring gravimeters which operate near room temperature and are more affected by environmental effects.

The levitation force of the sphere results in an interaction between the magnetic field from the superconducting coils and its induced current in the surface area of the sphere designed as hollow sphere with a small material thickness.

The levitation force can be adjusted by stepwise setting of the current (between 4 and 6A) in the two coils by the levitation electronic. It allows independent adjusting of both the levitation force and the magnetic gradient (corresponds to the spring constant in a mass-spring system). A weaker gradient gives a larger displacement of the sphere than a stronger one for the same gravity (acceleration) change. The SG

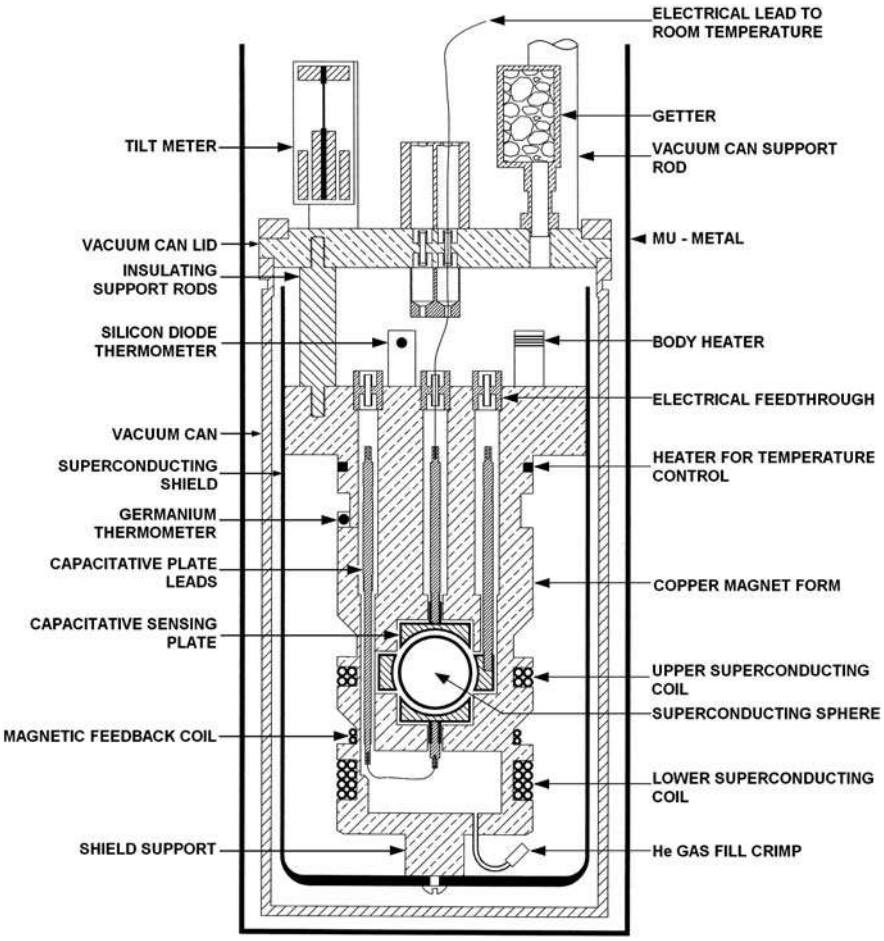


Fig. 10.3 Gravity sensing unit (drawing from GWR Instruments)

is designed with a weak restoring force to have relatively large displacements for small changes in gravity.

The displacement of the sphere which is proportional to gravity changes is measured by a capacitive transducer. The *capacitive sensing plates* of the transducer border the sphere with 1 mm clearance. The capacitive transducer in a Wheatstone bridge is part of a feedback system that holds the sphere in zero position relative to the gravity-sensing plates. A deviation from the sphere's zero position (capacitive bridge in equilibrium, null) caused by gravity changes generates a DC voltage proportional to the displacement (gravity change) that generates a current in the *magnetic feedback coils*. The resulting force brings the sphere back into zero position. The feedback system guarantees a linear transfer function between sphere displacement (gravity change) and the generated DC voltage for the entire

dynamic range of the sensor (~0.1 nGal–1.5 mGal). This DC voltage filtered by an anti-aliasing filter is input for the data acquisition system.

To avoid parasitic magnetic influences of the Earth's magnetic field and other sources on the superconducting elements the GSU is shielded with *MU-metal*. Additionally it has a *superconducting shield* and *insulating support rods*.

Because of the finite horizontal restoring force of the sphere, it can move in a horizontal direction too. Therefore, an orbital motion, an oscillation of the sphere, is generated by tilting the instrument. This oscillation appears as the sphere resonance with a period between 60 and 120 s depending on the instrument. The sphere resonance can be seen in the data, e.g. after local earthquakes (Imanishi 2005). For most of the SG applications this effect can be neglected because it can be filtered out.

10.2.2 Tilt Compensation System

The SG only measures the vertical component of gravity g_v exactly, if the vertical gravity sensor axis is aligned in the direction of the gravity vector. If there is a misalignment (angle θ) between both directions, the measured gravity g_m is $g_m = g_v(\cos\theta)$ and the tilt signal $g_T = g_v(1 - \cos\theta) \sim g_v \times \theta^2/2$. If the tilt signal g_T is smaller than 1 nGal, the tilt angle θ must be $\leq$1.4 μrad (Goodkind 1999). Therefore, the SG is equipped with an automatic tilt compensation system, which minimizes the tilt angle θ. It consists of two vertical sensitive pendulum tilt sensors orthogonally arranged (X and Y) within the gravity sensing unit of the SG (see Fig. 10.3). They are component parts in two separated feedback electronics which control the associated thermo-mechanical levellers X and Y (Fig. 10.2). The SG has three dewar support posts, a fixed one and the two adjustable X and Y, which correspond to tilt in *X* and *Y* directions (tiltX and tiltY). In the SG setup procedure the tilt angle is preset to a minimum by the two micrometer screws (Fig. 10.2). During SG operation the automatic tilt compensation system holds the tilt angle at a minimum by aligning the instrument (dewar) with the two thermo-mechanical levellers, which are unfortunately temperature sensitive. Therefore, the room temperature of the SG should be stable within the range smaller than ±1.5°C. Checking of the tilt can be performed by the tilt (X and Y) power and balance signals which are proportional to the tilt angle. They are recorded by the data acquisition system of the SG. The tilt compensation system corrects the tilt signal g_T to about 1 nGal, which corresponds to about ±5 mV variation in the tilt balance signals.

10.2.3 Dewar and Compressor

The operation of the GSU at liquid helium temperature requires a refrigeration system that compensates heat conduction between the operation temperature of 4.2 K and room temperature. It must compensate the heat conduction of the dewar which is small compared to those caused by the GSU (less 0.1 mW). This heat energy leads to evaporation and a loss of the liquid helium. A compressor and a coldhead, mounted

in the neck of the dewar, reduce the consumption of liquid helium. Presently for the SG the compressor CAN 11 and the coldhead RDK-101 from Sumitomo Heavy Industries (SHI) are used. The coldhead has two temperature stages around 60 and 4 K. It produces a cooling power of about 0.2 W at the 4 K (lower stage). This is a closed-cycle cryogenic system where the helium gas condenses in the neck at the lower stage and drips back into the dewar (Richter and Warburton 1989). It can liquefy about 1 L/day. The compressor needs a power of 1.3 kW and it is air cooled.

The compressor including the coldhead causes vibrations of about 2 Hz up to hundreds of Hertz. These vibrations must not bias the dewar with the GSU. Consequently the coldhead is suspended on a separate frame which is vibration insulated from the floor of the SG pillar. The coldhead is arranged in such a way that it does not touch the neck of the dewar.

Different dewar sizes were in use in the past (Hinderer et al. 2007). Presently the most manufactured dewar for the SG has a volume of 35 L for the liquid helium. This volume of liquid helium allows an operation of about 20 days in case of failure of the compressor or coldhead.

10.2.4 Gravimeter Electronic Package

The gravimeter electronic package (GEP) consists of the control electronics for gravity, temperature and tilt; the data acquisitions system; and a current supply for the sphere levitation. The data acquisition system records gravity, the SG operation parameters and the environmental parameters, e.g. atmospheric pressure.

For gravity recording the analogue anti-aliasing filter is designed for a 1-s sampling rate (GGP filter). It has a damping of – 160 dB/decade, a corner frequency of 61.5 mHz and a time delay of 8.4 s. It is in use for all SGs as standard. The analogue gravity signal with a dynamic range of ± 10 V is digitized by a digital voltmeter with a resolution of 7.5 digits, which corresponds to the smallest significant voltage change of 0.3 μV. With the ratio of 100 μGal/V (calibration coefficient around 70 μGal/V) the significant change of gravity is 0.03 nGal. This is the resolution (quantization) limit of the data acquisition system. It is above the SG's gravity resolution.

The data acquisition system includes the remote control capability of the instrument. It allows setting of all control parameters for gravity, temperature and tilt and the data acquisition parameters. Moreover, the levitation of the sphere can be carried out remotely.

10.2.5 SG Performance

The SG is characterized by a resolution of about 0.1 nGal in the frequency domain. The precision can be estimated only by means of analysed known gravity effects, because there is no reference gravity better than from the SG itself. The precision is frequency dependent and differs in frequency and time domain. In frequency domain

gravity signals of 1 nGal and better can be detected in the band from the free oscillation of the Earth to the diurnal Earth tides (~10–0.01 mHz). In time domain the precision is better than 0.05 μGal depending on the noise level at the site (Hinderer et al. 2007).

The SG recordings cover a period range from minutes to years with a linear transfer function (constant gain = 1). The drift behaviour shows an initial exponential drift followed by a linear drift rate of some microgalileos per year depending on the instrument.

The dual sphere SG has the same parameters for each sensor as the single sphere SG. Additionally the dual sphere system measures the gravity gradient with a resolution of about 0.5 μGal/m (5 Eötvös).

More details about the behaviour of the dual sphere instrument can be found in Kroner et al. (2005) and more design details of the instrument in Goodkind (1999).

In Table 10.2 the main SG performance parameters are summarized.

Table 10.2 Summary of the main SG performance parameters

Resolution	0.1 nGal
Precision in frequency domain	<1 nGal
Precision in time domain	<0.05 μGal
Period range	Minutes to years
Measurement range	~0.1 nGal–1.5 mGal
Accuracy calibration factor	~±0.05 μGal/V (~±0.2%)
Gravity time shift (standard)	8.4 s
Gravity filter corner frequency (standard)	61.5 mHz
Drift rate	Some microgalileos per year

10.3 Site Selection and Observatory Design

The site selection depends first on the scientific goals the SG shall be used for. According to these goals (cf. Sect. 10.1), e.g. investigation of global surface gravity effects, local and regional hydrology, silent earthquakes, tectonic movements, those sites should be selected where the largest signal of the gravity effect of interest is expected. Furthermore, for the site selection some other aspects have to be considered, which can corrupt the recorded gravity signal by overlaying with a disturbing signal:

- Industrially induced seismicity (traffic, industry, mining activities, etc.)
- Natural seismicity (microseismic activities, trees in windy areas, etc.)
- Earthquakes
- Ocean noise close to the coast
- Uncontrolled groundwater level changes
- Bad coupling of the pillar to the ground
- Tectonic motions

Therefore, before constructing the observatory the disturbing signals should be estimated by studying the

- Noise, recorded with a broadband seismometer
- Geological situation (geological province and characteristics, depth of bedrock)
- Hydrological situation (groundwater variations, water pumping stations, lakes and rivers in the surrounding)

The knowledge of these factors gives the first estimation of the disturbing signal in the gravity data and this knowledge can be very helpful in fixing the location of the site. Of course, the site selection depends mostly on the scientific goals, but the disturbing signal must be smaller than the expected signal of interest.

The SG requires a stable well-grounded platform and a temperature-controlled measurement chamber. The SG fits comfortably on a concrete (or granite) pillar of 80 cm $\times$ 80 cm. The electronics and data acquisition system are fitted into one small electronic rack, which is temperature controlled. This rack is a standard 19" width rack. It is 28" high and 24" deep. The minimum overhead clearance to transfer liquid helium is 180 cm. The cables connecting the GEP electronics to the dewar head are about 2.5 m long. They must, however, be routed loosely between the electronics and dewar in a way so that they do not touch the cold-head or coldhead frame. This usually requires that the electronic rack should be placed between 0.5 and 1 m away from the dewar. The power requirement is at least about 2 kW.

For the SG installation only one pillar of about 80 $\times$ 80 cm is necessary. But it is to be considered that absolute gravimeter measurements (e.g. with FG5) are necessary for exact calibration of the SG. Therefore, it is recommended to build a second pillar (about 90 $\times$ 90 cm) for the absolute measurements or enlarge the pillar to about 200 $\times$ 100 cm. To avoid disturbances on the SG, a second pillar should be the better solution.

For housing of the SG, a small building or container with two rooms is sufficient: one room for the SG (measuring chamber) and the second one for the SG electronics, compressor and air-conditioning equipment.

The measuring chamber contains the SG pillar and the pillar for absolute measurements. The SG room should be air conditioned and the temperature should be stable within about ± 1.5°C. The room temperature depends on the location of the SG site. Presently the SG sites have a working temperature for the SG below of about 25°C. The humidity in the room should be stable within $\pm 10\%$. To minimize temperature variations, the measuring chamber should be covered with insulating material. For reducing environmental influences like temperature changes and strong wind that causes noise, the measuring chamber can also be built into the ground.

Figure 10.4 shows a design example for a SG observatory layout similar to the SG station in Sutherland (SU) in South Africa (32.3814°S, 20.8109°E and height 1,791 m).

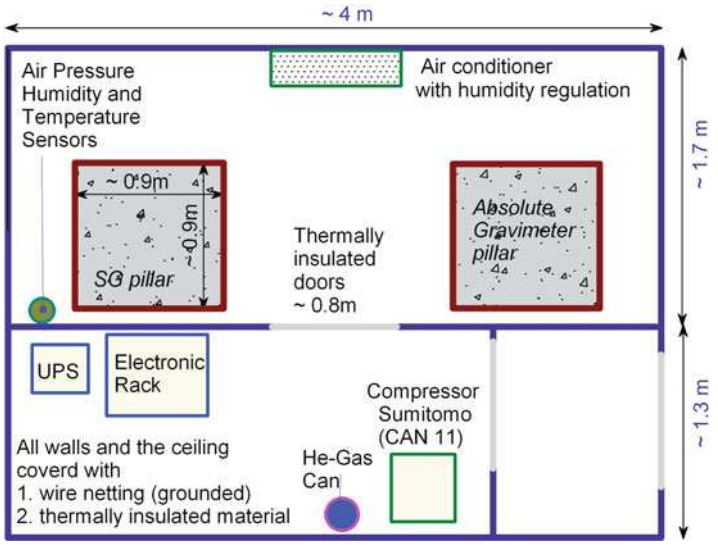


Fig. 10.4 Design example observatory layout

The compressor is placed a few metres away from the SG to avoid noise in the SG data. For avoiding this noise the compressor can stand on a damping material like a rubber plate.

The pillar is the coupling element between the SG and the Earth. Therefore, it must be constructed very stably, e.g., of concrete with reinforcement and with three or four drill holes (Fig. 10.5). If bedrock is available, the pillar should be placed on it. It is not recommended to build the pillar on a floating ground. The pillar should be displaced from the floor of the measuring chamber to avoid noise coming from the floor and the observatory building.

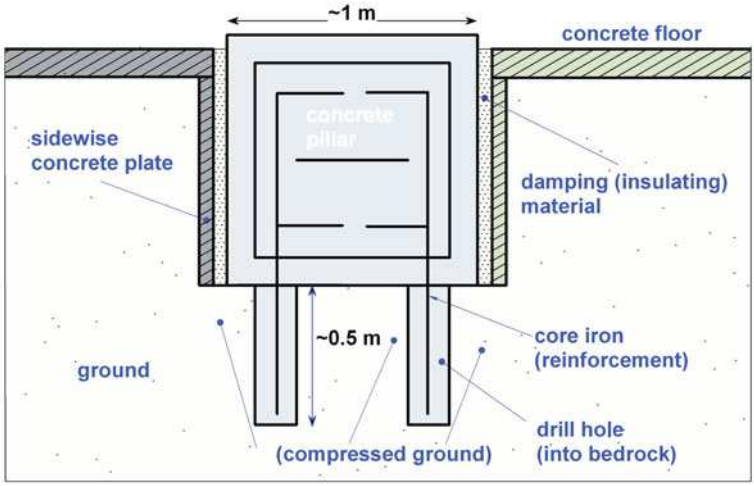


Fig. 10.5 Design example (section) of the measurement pillar

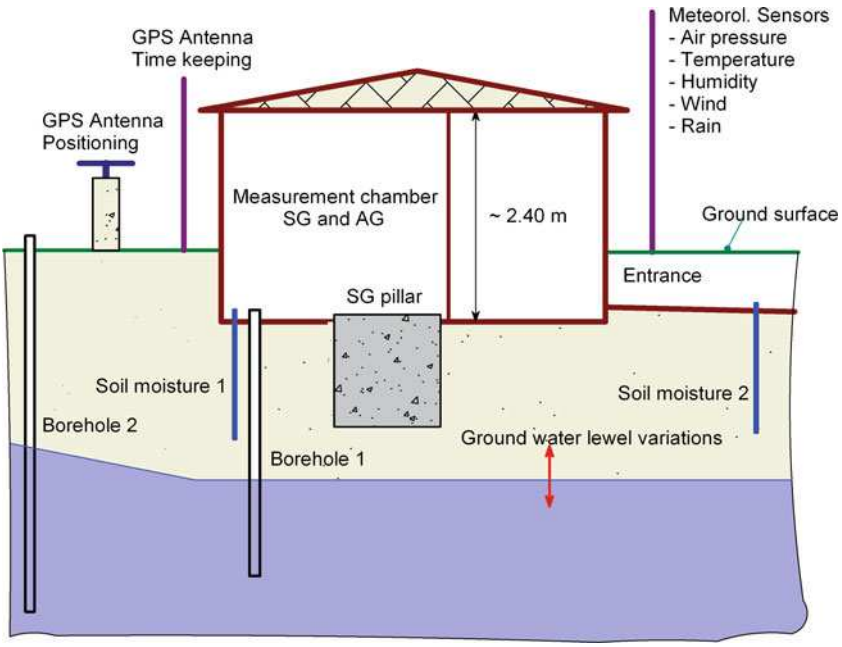


Fig. 10.6 Design example SG observatory

Figure 10.6 shows a design example of a SG observatory with the arrangement of the meteorological and hydrological sensors, also including GPS antennas for time keeping and positioning. The location of borehole 2, 3, etc. and soil moisture sensors 2, 3, 4, etc. depends on the hydrological conditions and the topography at the site (cf. Sect. 10.6.3).

10.4 Calibration of the Gravity Sensor

The data acquisition system records the voltage of the feedback system of the gravity sensor, which is proportional to the gravity changes. Therefore, the collected raw gravity data must be calibrated in amplitude by multiplication with the calibration (scaling) factor. It has to be determined individually for each instrument. Besides the amplitude calibration, the recorded SG signal has a phase delay caused by the gravity sensor and its electronic including the analogue anti-aliasing filter and the analogue to digital converter (ADC). This phase delay also has to be determined and the recorded gravity data must be corrected in time.

10.4.1 Calibration Factor

The state of the art in calibration of a SG is the derivation of the calibration factor from a parallel gravity registration of SG and absolute gravimeter (e.g. FG5) or

a well-calibrated spring gravimeter (e.g. LaCoste and Romberg or Scintrex). The calibration experiment should be carried out for at least about 5 days.

For determination of the calibration coefficient, the data should be corrected for outliers (e.g. 3σ) and big disturbances, e.g. earthquakes should be extracted from both data sets. By a linear regression of parallel recorded SG and absolute gravimeter data sets, the calibration factor can be estimated with an accuracy of about ± 0.1 μGal/V, which corresponds to about $\pm 0.15\%$ (Hinderer et al. 1998; Francis and van Dam 2002). With the inertial calibration platform (BKG Frankfurt, Germany), an accuracy of ± 0.02 μGal/V could be achieved (Falk et al. 2001).

A theoretical approach for determination of the scaling (calibration) factor is based on theoretical Earth tides as reference. The determination of the scaling factor is performed by linear regression between the raw SG gravity data and the theoretical Earth tides at the SG location, based on the Hartmann–Wenzel tidal catalogue HW95, which has an accuracy of 1 nGal, (Hartman and Wenzel 1995a, b) and the Wahr–Dehant Earth tide model (WD model) (Dehant 1987) by using, e.g., the program PREDICT (Wenzel 1996). Gravity variations caused by the atmosphere, the hydrosphere, the ocean and the polar motions, according to Sects. 10.6.2–10.6.5, should be added to the theoretical Earth tides, because the raw gravity data also include these signals.

Figure 10.7 shows as example the calibration of the dual sphere SG D037 at Sutherland station in South Africa (Neumeyer et al. 2002). Three different methods were applied:

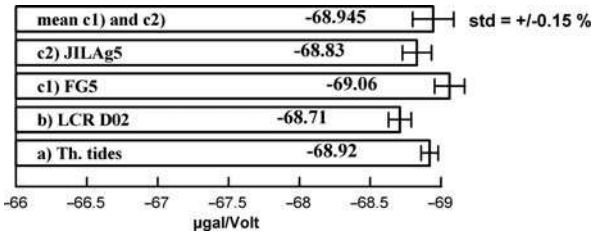


Fig. 10.7 Calibration results for SG D037 (lower sensor)

1. The first calibration factor was determined by linear regression between the theoretical tides added to gravity induced by the atmosphere (single admittance, cf. section "Single Admittance"), polar motion (cf. Sect. 10.6.5) and the ocean tides (cf. Sect. 10.6.4.2) for the SU site and the output gravity signal of the SG for a time from July to December 2000.
2. Parallel registration of the SG and the LaCoste and Romberg feedback gravimeter D02 which has been calibrated at the Hanover calibration line (Rehen 1997). The calibration factor was determined by comparison of the Earth tide analysis results of both data sets. The SG calibration factor was varied in the SG Earth tide analysis until the amplitude factor for the partial tide O1 had the same value

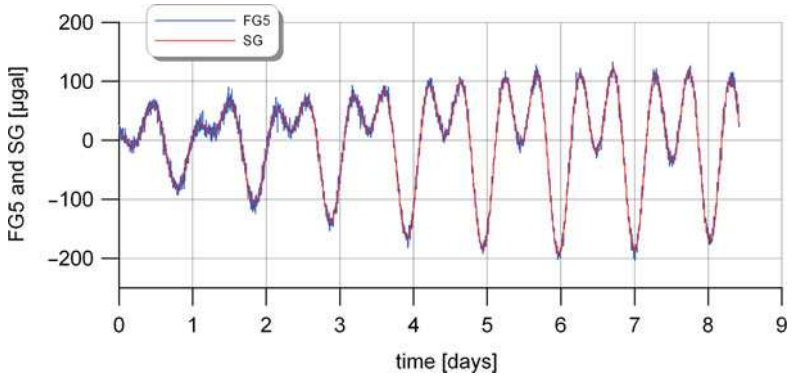


Fig. 10.8 Recorded SG (*red*) and FG5 (*blue*) data

as the O1 amplitude factor determined by the LaCoste and Romberg Gravimeter Earth tide analysis (reference).

3. Parallel registration of the SG and an absolute gravimeter. In February and March 2001 a parallel registration was carried out with the absolute gravimeters:

   a. FG5 from "Ecole et Observatoire des Sciences de la Terre", Strasbourg, France, from 1 to 9 February 2001
   b. JILAg5 from "Finnish Geodetic Institute", Masala, Finland, from 21 to 29, March 2001

From both time series the absolute gravimeter and the raw SG data outliers larger than 3σ and the linear trend were removed. Figure 10.8 shows the parallel recorded SG and FG5 data used for determination of the calibration factor. It was performed by a linear least square fit between the absolute gravimeter and the SG data. Figure 10.9 shows the plot of FG5 (Y) vs. SG (X) data $Y = f(X)$ and the regression line $r(X) = b_0 + b_1 X$. The factor b_1 corresponds to the calibration coefficient.

As the final calibration factor the mean of the parallel registrations of the absolute gravimeters FG5 (a), JILAg5 (b) and the SG, $CF_{ag} = -68.945$ µGal/V was selected. The calibration factor was determined with a standard deviation of $\pm 0.15\%$.

Remarkable is the difference between the two absolute gravimeter results of 0.77 µGal/V (0.5%) and the good agreement between the calibration factors based on theoretical tides and the mean of the two absolute gravity measurements. The difference between the scaling factor based on theoretical tides, $CF_{Th} = -68.92$ µGal/V and CF_{ag} is 0.025 µGal/V, which lies within the error bar of the measured calibration coefficient CF_{ag}. This result shows that the scaling factor determination based on theoretical tides is in good agreement with the calibration based on absolute gravimeter recordings.

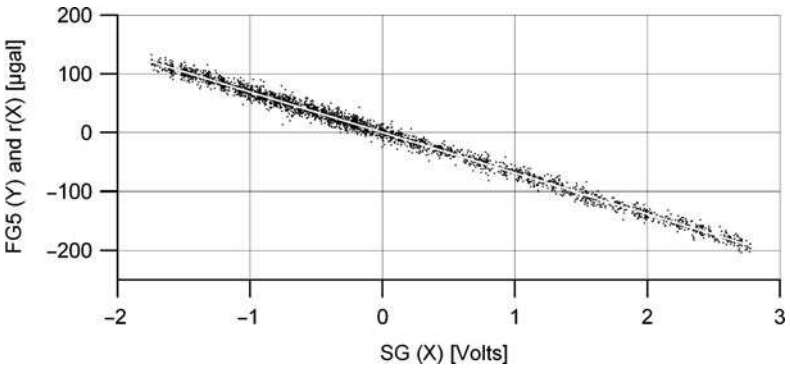


Fig. 10.9 Regression between FG5 and SG data (*black*) and regression line (*white*)

10.4.2 Phase Shift

The phase shift is the delay of the recorded gravity signal with respect to original time of this signal. This time delay is caused by the gravity sensor and its electronics including the anti-aliasing filter, which has a nominal time delay of 8.4 s, and the analogue to digital converter. The determination of the phase shift is necessary for the exact time assignment of the gravity signal.

The phase shift determination is based on an additional step or sine voltage in the feedback coil. This voltage corresponds to a gravity signal. It can be added by an adder circuit incorporated into the feedback circuitry. In this way the complex transfer function and the time delay of the SG can be determined (Richter and Wenzel 1991; Wenzel 1995a; van Camp et al. 2000).

For a step voltage as input signal the complex transfer function of the SG, $\hat{H}(\omega) = \mathrm{Re}(H(\omega)) + \mathrm{I}m(H(\omega))$, can be determined by Fourier transform of the differentiated recorded step response voltage (Bendat and Piersol 1986). From $\hat{H}(\omega)$, the amplitude transfer function (gain) $G(\omega)$ and the phase shift $\varphi(\omega)$ can be derived according to (1) and (2):

$$G(\omega) = \sqrt{[\mathrm{Re}(H(\omega))]^2 + [\mathrm{Im}(H(\omega))]^2}, \tag{1}$$

$$\varphi(\omega) = \arctan \frac{\mathrm{Im}(H(\omega))}{\mathrm{Re}(H(\omega))}. \tag{2}$$

The frequency-dependent time shift $\mathrm{ts}(\omega)$ can be determined by (3) with $T =$ period of the relevant frequency range, e.g. the Earth tides:

$$\mathrm{ts}(\omega) = \frac{\varphi(\omega)}{360^\circ} T \tag{3}$$

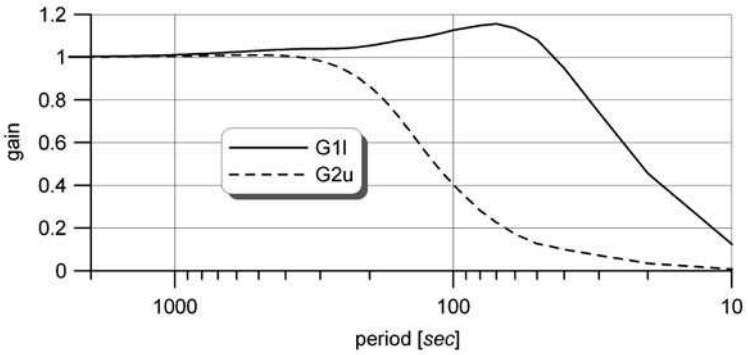


Fig. 10.10 Amplitude transfer function of SG D037 (G1l = lower and G2u = upper sensor)

An example for the determination of the transfer function and time delay is given for SG D037 at SU station. The signal added to the feedback circuitry was generated by a low-frequency function generator. The new electronic package GEP includes a function generator for a step and sine wave experiment to determine the complex transfer function and time shift of the SG.

In a step response experiment the time delay of the lower sensor was determined as ts = 8.7 s. The calculation of the amplitude transfer function $G(\omega)$ and the phase shift φ can be carried out, e.g., by the programs ETSTEP (Wenzel 1996) or TSOFT (Van Camp and Vauterin 2005). The phase shift can be determined with an accuracy of about 0.01°.

Figure 10.10 shows the amplitude transfer function for the lower and upper sensors of SG D037. The different transfer functions result in the different gradients (lower sensor (G1l) weak and upper sensor (G2U) stiff gradient) which were generated during the setup procedure of the SG. The gradient corresponds to the spring constant of a mass-spring gravimeter.

10.5 Noise Characteristics

The investigation of weak gravity effects requires a high signal to noise ratio, which characterizes the quality of the data. The signal to noise ratio of the recorded gravity data depends on the noise at the site and the noise of the instrument. Because the instrumental noise is small in the inspected frequency band, the noise level at the site mostly determines the quality of the data.

10.5.1 Noise Magnitude

To estimate the noise level at a SG site in the high-frequency range, the noise magnitude (NM) is often used (Banka and Crossley 1999). The NM can be calculated by

$$\mathrm{NM} = 10\ \log(\mathrm{PSD})\ (\text{related to}\frac{\mathrm{m/s}^2}{\sqrt{\mathrm{Hz}}}, \text{in dB}). \tag{4}$$

The power spectral density (PSD) is normally calculated from 1 s sampled raw gravity data, corrected for air pressure-induced gravity. The quality of the SG site can be estimated by comparison of the calculated noise magnitude with the New Low Noise Model (Peterson 1993). Rosat et al. (2004) compared the noise magnitude of 19 SG sites of the GGP network. Because of the different locations and environment conditions at the sites the noise magnitudes differ up to about 20 dB.

As an example, the noise magnitudes were calculated for the SG stations in SU in South Africa (one of the low-noise stations according to Rosat et al. 2004) and MunGyung (MG) site in South Korea (36.6402°N, 128.2147°E and height 107.5 m). The used data were for the SU site from 15 to 31 May 2005 (lower gravity sensor), and for the MG site from 1 to 16 January 2007.

The PSD was computed with a subdivision length of the raw gravity data of 12 h and an overlapping of 50%, and the data were multiplied with a Hanning window. The loss of energy in using the Hanning window was compensated by multiplication of the Fourier components in the PSD computation with a factor of $\sqrt{8/3}$ (Bendat and Piersol 1986).

In Fig. 10.11 the PSD (left axis) and the associated NM (right axis) are shown as functions of frequency in millihertz. For comparison the NM, according to the NLNM, is added.

A comparison between the noise magnitudes of the SG sites and the NLNM shows that the NM of the SU site is close to the NLNM below 10 mHz and even a little smaller below 1 mHz, whereas the NM at the MG site is higher than at the SU site within this frequency range. Above 10 mHz the NM at the MG site is smaller.

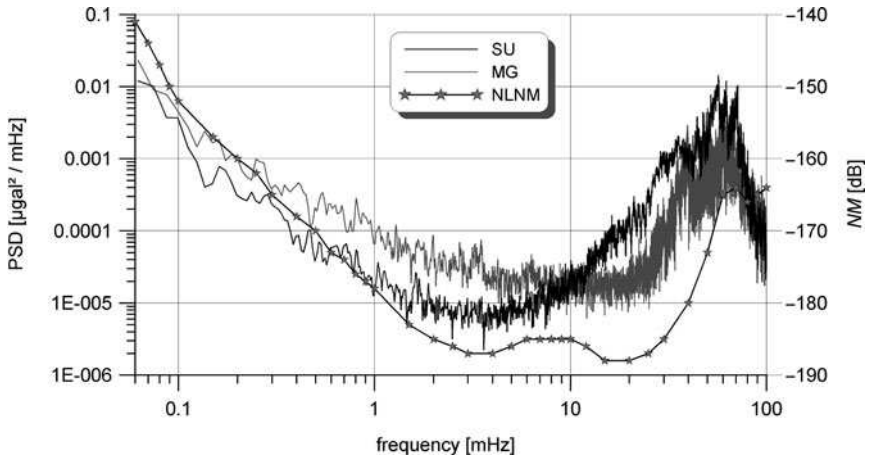


Fig. 10.11 Power spectral density (PSD) and noise magnitude (NM) at SU and MG sites in comparison with the New Low Noise Model (NLNM)

The peak of the NM in the range from about 30 to 100 mHz is mainly caused by the ocean noise. This frequency range is normally filtered out because the interesting gravity signals of the SG observations begin at frequencies below 10 mHz (free oscillations of the Earth).

This comparison with the NLNM model and the results of the other analysed SG sites by Rosat et al. (2004) emphasize that the SG recordings on a quiet site offer excellent conditions for high-precision gravity measurements and the detection of weak gravity signals in the frequency range below 10 mHz, e.g. for the free oscillation of the Earth and the translational oscillation of the solid Earth core.

10.5.2 Noise Caused by Misaligned Instrumental Tilt

If the tilt compensation system does not work properly, the tilt changes cause noise in the gravity data. This behaviour was observed by Imanishi (2005) and Kim et al. (2009).

The influence of the misaligned tilt on gravity recordings is shown for the MG station in Korea (Kim et al. 2009). Since the automatic tilt system of the SG at MG site was found not to work properly, it was newly adjusted and the air blowing of the air-conditioning system to the thermo-mechanical levellers was drastically reduced on 14 June 2006 (Event #1). In addition, the SG was covered with a polystyrene box to maintain the ambient temperature more stable on 21 September 2006 (Event #2). Figure 10.12 shows the recording of tiltY power and balance signals before and after Events #1 and #2.

There is a small decrease of tilt power variation after Event #1 and a decrease from about ±50 to ±20 mV after Event #2. The tilt balance signal comes up to about ±25 mV before Event #1 and after it was reduced to about ±10 mV. A further decrease of the balance signal could be achieved by box installation (Event #2), which results in a medium tilt balance signal of ±5 mV. This is the nominal value

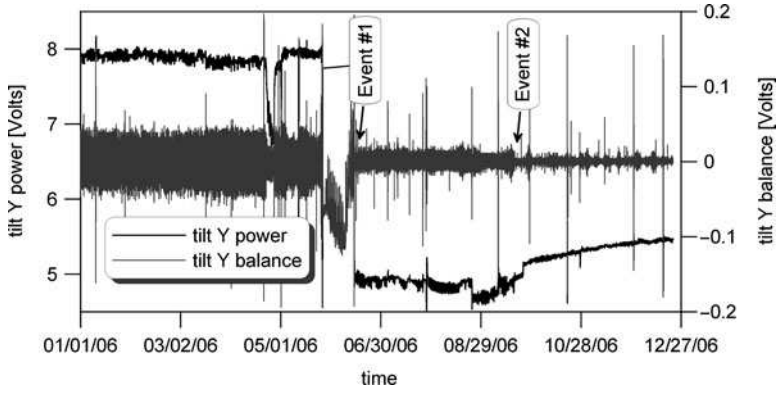


Fig. 10.12 TiltY power and balance recordings

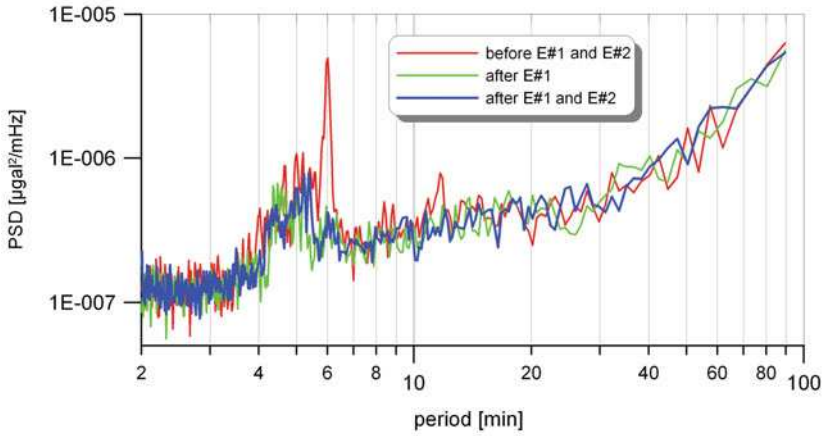


Fig. 10.13 Power spectral density of the gravity residuals δg_{res} before and after Events #1 and #2

for a correctly working tilt compensation system. TiltX power and balance signals showed very similar behaviour like that of tiltY.

The influence of the misaligned tilt on gravity recordings is shown in Fig. 10.13. It displays the PSD of the gravity residuals ($\delta g_{\mathrm{res}} = \delta g_{\mathrm{raw}} - \mathrm{ET} - \delta g_{\mathrm{air}} - \delta g_{\mathrm{ol}} - \delta g_{\mathrm{gwl}}$) (cf. Sect. 10.7.7) 7 days before and after Events #1 and #2. The PSD was calculated with a subdivision length of the gravity residuals of 7 h, an overlapping of 50% and the data were multiplied with a Hanning window. There are peaks at the same periods (between 4 and 7, 8 and 9, and around 12 min), as in the spectrum of tilt balance calculated previously. These peaks became smaller or disappeared after Event #1. After Event #2, only a remarkable reduction of the peaks between 4 and 5 min was achieved. This result confirms that the tilt balance signal should be ±5 mV or smaller to avoid disturbances in the gravity signal.

10.6 Modelling of the Principal Constituents of the Gravity Signal

The SG measures gravity variations according to Newton's law of gravitation resulting from mass movements in its near and far surroundings. Therefore, the SG signal includes gravity effects from the Moon and Sun as well as from other celestial bodies, the Earth tides and gravity effects induced by various geophysical and geodynamic sources within the Earth's system on global, regional and local scales. Figure 10.14 shows a typical example of the gravity signal recorded with the SG at station Metsahovi (ME), Finland (60.2172°N, 24.3958°E and height 55.6 m), from April 2002 to July 2003 and the modelled principal constituents of the gravity signal:

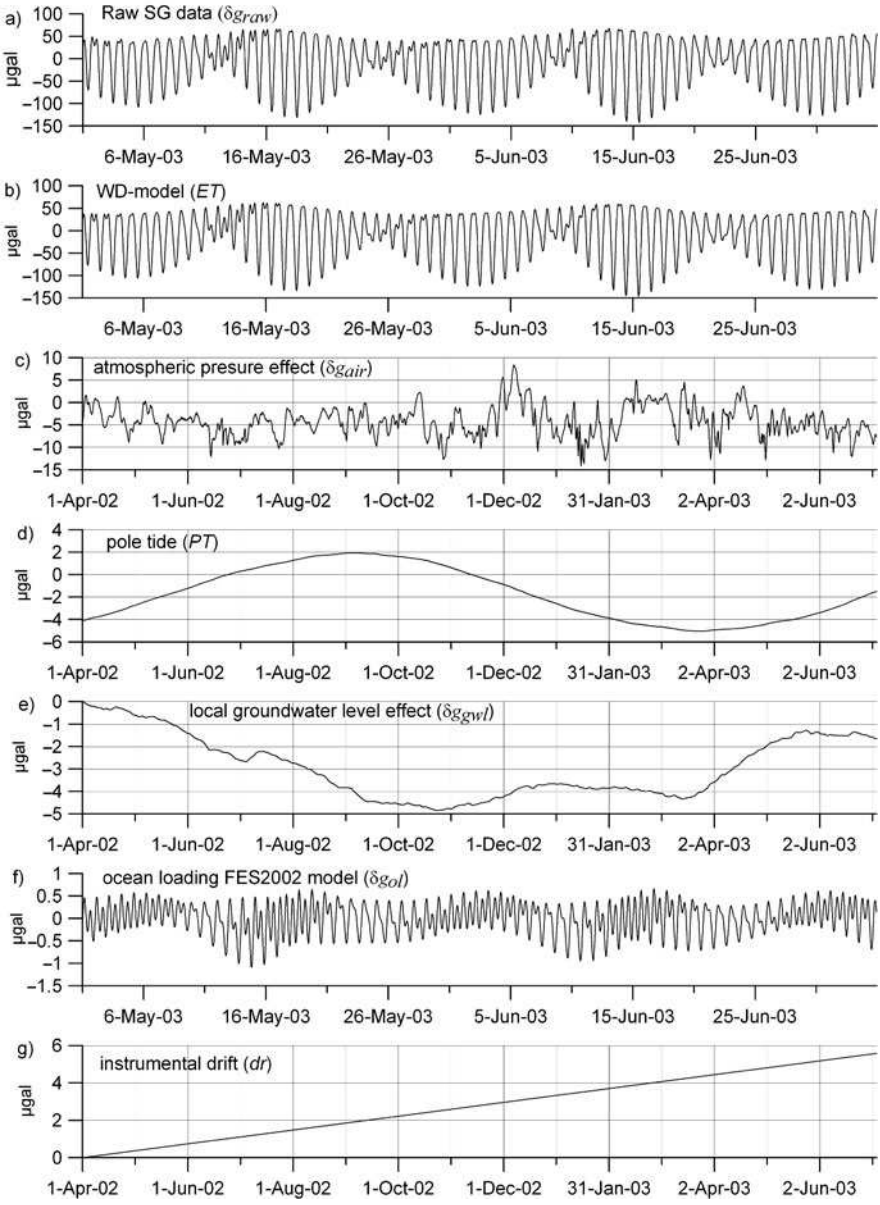


Fig. 10.14 Gravity signal and its principal constituents (station Metsahovi, Finland): (**a**) raw gravity data (δg_{raw}); (**b**) theoretical Earth tides due to WD model (ET); (**c**) atmospheric pressure effect (δg_{air}); (**d**) pole tide based on IERS data (PT); (**e**) local groundwater level effect (δg_{gwl}); (**f**) ocean loading based on FES2002 model (δg_{ol}) and (**g**) instrumental drift (dr)

1. Recorded raw gravity data (δg_{raw})
2. Theoretical Earth tides due to the WD model (ET)
3. Atmospheric pressure effect (δg_{air})
4. Pole tide (PT) due to IERS data
5. Local groundwater level effect (δg_{gwl})
6. Ocean loading due to FES2002 model (δg_{ol})
7. Instrumental drift (dr)

The recorded raw gravity which includes all gravity signals has peak to peak amplitude of 213.12 μGal (Fig. 10.14a). The largest signal in the raw gravity data is caused by the Earth tides (ET). They are calculated according to the HW95 tidal catalogue (Hartmann and Wenzel 1995a, b) and WD model (Dehant 1987) with the ETERNA package (Wenzel 1996) shown in Fig. 10.14b. The ET signal has peak to peak amplitude of 210.08 μGal. Large signals are induced by the mass redistribution in the atmosphere (Fig. 10.14c). In this example the gravity variations are 22.45 μGal calculated with a single admittance coefficient of –0.313 μGal/hPa as described in section "Single Admittance". The hydrologically induced gravity effect caused by groundwater level changes in the surrounding of the SG (Fig. 10.14e) comes to 4.85 μGal. It was calculated with an admittance coefficient of 3.3 μGal/m (cf. Sect. 10.6.3.3). The polar motion-induced gravity effect (Fig. 10.14d) calculated according to Sect. 10.6.5 reaches 7.0 μGal. Figure 10.14f shows the ocean tide loading effect with peak to peak amplitude of 1.94 μGal calculated according to Sect. 10.6.4. Not negligible is the instrumental drift of 4.2 μGal/year (cf. Sect. 10.6.6).

10.6.1 Theoretical Earth Tides and Tidal Acceleration

The changing gravitational forces from Sun, Moon and the planets affect the gravity acceleration g at the Earth. In the Earth's centre of mass this gravitational force F_{gr} is compensated by the centrifugal force F_{c} due to the motion of the Earth around the Sun and due to the motion of the Earth around the barycentre (BC) of the Earth–Moon and Earth–planet system, respectively, which is illustrated for a two-body system (e.g. Earth–Moon) in Fig. 10.15. The centrifugal acceleration (a_{c}) is constant in every point of the Earth but the gravitational acceleration (a_{gr}) is different due to the spatial extent of the Earth. The small resulting acceleration $\vec{a}_{\text{tid}}$ is called tidal acceleration which causes an elastic deformation of the Earth, the Earth tides.

The calculated tidal acceleration $\vec{a}_{\text{tid}}$ can be determined at point P according to (5) with $\vec{a}_0$ = gravitational acceleration at the geocentre, GM = gravitational constant times mass of the celestial body:

$$\vec{a}_{\text{tid}} = \vec{a}_{\text{P}} - \vec{a}_0 = \text{GM}\left(\frac{d}{d^3} - \frac{s}{s^3}\right) \tag{5}$$

As we see in Fig. 10.15, the tidal acceleration $\vec{a}_{\text{tid}}$ is distributed symmetrically; therefore no additional acceleration of the whole Earth is generated. Because of the

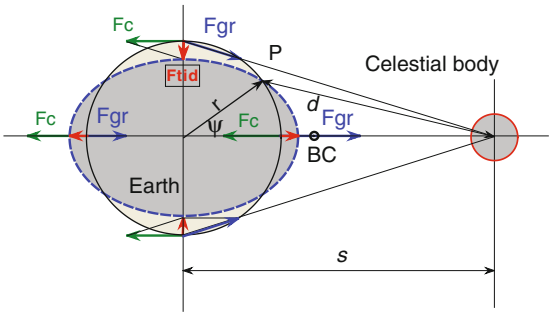


Fig. 10.15 Earth tides. F_{gr} = gravitational force of the celestial body (Sun, Moon, planets), F_c = centrifugal force, F_{tid} = tidal force, BC = barycentre, s = distance between mass centres of the Earth and the celestial body, d = distance of the mass centre of the celestial body to point P at the Earth's surface, ψ = geocentric zenith angle of Moon, Sun or a planet, r = geocentric distance of the attracted point (on the Earth's surface $r \sim R$)

declination of the Earth rotation axis Ω (ecliptic) the tidal acceleration is changing at any position of the Earth with time. The largest Earth tides have diurnal and semidiurnal periods.

For calculation and prediction of the theoretical Earth tides the tidal potential has been introduced based on spherical harmonic expansion of the tidal potential. The tidal acceleration vector $\vec{a}_{tid}$ is the gradient of the tidal potential V:

$$\vec{a}_{tid} = \text{grad } V = \frac{\partial V}{\partial \vec{r}}. \tag{6}$$

The solution for this differential equation results in (7) for $V = 0$ in the geocentre by using the Legendre polynomial expansion $P_\ell \cos(\psi)$:

$$V = \frac{GM}{s} \sum_{\ell=2}^{\infty} \left(\frac{r}{s}\right)^{\ell} P_\ell(\cos\ \psi). \tag{7}$$

Because of the fast convergence of the spherical harmonic expansion, degree $\ell = 6$ for the Moon, $\ell = 3$ for the Sun and $\ell = 2$ for the planets are sufficient. For separation of the latitude and time-dependent parts of the tidal potential, the geocentric zenith angle ψ is expressed by the spherical coordinates of point P (see Fig. 10.15) and the celestial body. Additionally, the Legendre polynomials $P_\ell \cos(\psi)$ are decomposed in normalized spherical harmonics $\overline{P}_{\ell,m}$ of degree ℓ and order m:

$$V(t) = GM_{Pl} \sum_{\ell=2}^{\infty} \sum_{m=0}^{\ell} \frac{r^\ell}{s^{\ell+1}} \frac{1}{(2\ell+1)} \overline{P}_{\ell,m}(\cos\ \theta) \overline{P}_{\ell,m}(\cos\ \Theta_{Pl}(t))\ \cos(m(\lambda - \Delta_{Pl}(t))) \tag{8}$$

with the geocentric spherical coordinates θ, λ at point P, and Θ, Λ at the centre of mass of the celestial body (Wenzel 1995b).

For the Hartmann–Wenzel tidal catalogue HW95 (Hartmann and Wenzel 1995b), (9) has been used for calculation of the tidal potential $V(t)$:

$$V(t) = \sum_{\ell=1}^{\ell\,\max} \sum_{m=0}^{\ell} \left(\frac{r}{a}\right)^{\ell} \overline{P}_{\ell,m}(\cos\,\theta) \sum_i \left[C_i^{\ell,m}(t)\cos(\alpha_i(t)) + S_i^{\ell,m}(t)(\sin(\alpha_i(t)) \right] \quad (9)$$

with the major semiaxis of the Earth $a = 6378136.3$ m and time t from J2000 in Julian centuries. The time-dependent tidal potential coefficients are given by

$$C_i^{\ell,m}(t) = C0_i^{\ell,m} + t\, C1_i^{\ell,m}, \quad (10)$$

$$S_i^{\ell,m}(t) = S0_i^{\ell,m} + t\, S1_i^{\ell,m}. \quad (11)$$

The arguments $\alpha_i(t)$ are computed from

$$\alpha_i(t) = m\,\lambda + \sum_{j=1}^{11} k_{i,j}\ \mathrm{arg}_j(t) \text{ with } \mathrm{k}_{i1} = m. \quad (12)$$

The potential coefficients $C_i{}^{\ell,m}$, $S_i{}^{\ell,m}$ and the integer coefficients k are given in the HW95 catalogue, while the 11 arguments $\mathrm{arg}_i(t)$ can be computed from polynomials in time, given in Hartman and Wenzel (1995b).

The different components of the tidal acceleration $\vec{a}_{\mathrm{tid}}$ can be calculated by partial differentiation of the tidal potential with respect to the spherical coordinates (r,θ,λ). The radial component of the tidal acceleration $a_{\mathrm{tid}\text{-}r}$ can be calculated from (13) for any frequency (tidal wave) contained in the catalogue (e.g. $i = 12{,}935$ (HW95) and $i = 1{,}200$ (Tamura 1987) tidal waves):

$$a_{tid-r} = \frac{\partial V}{\partial r} = \sum_{\ell=1}^{\ell_{\max}} \ell \frac{r^{\ell-1}}{a^{\ell}} \sum_{m=0}^{\ell} \overline{P}_{\ell,m}\,(\cos\,\theta) \sum_i \left[C_i^{\ell,m}(t)\cos(\alpha_i(t)) + S_i^{\ell,m}(t)\sin(\alpha_i(t)) \right]. \quad (13)$$

Because the tidal acceleration is needed for an ellipsoidal coordinate system of the Earth, the angular difference β between the ellipsoidal and the geocentric latitude must be considered according to

$$a_{\mathrm{Etid}-r} = \sin(\beta)\, a_{\mathrm{tid}-r} + \cos(\beta)\, a_{\mathrm{tid}-r}. \quad (14)$$

The changes of the tidal acceleration $a_{\mathrm{Etid}\text{-}r}$ cause a deformation of the Earth, the Earth body tides. This deformation causes an altitude change combined with a mass movement at a fixed point P at the Earth's surface. The Earth tides cannot be measured directly. However, they can be determined by the response of the Earth to the tidal acceleration. This response can be measured with a gravimeter in amplitude

and phase and compared to the calculated theoretical tidal acceleration $a_{\text{Etid-}r}$. The deviation of the amplitude and phase of a partial Earth tide wave ET_i at point P of the Earth's surface due to altitude change and mass movement compared to the theoretical tidal acceleration $a_{\text{Etid-}r}$ can be expressed by

$$ET_i(t) = \delta_i \, A_{\text{th}_i} \, \cos(\omega \, t + \delta\Phi_i) \tag{15}$$

with A_{th_i} = amplitude of theoretical tidal acceleration $a_{\text{Etid-}r}$, $\delta\Phi_i = \Phi_{\text{obs}_i} - \Phi_{\text{th}_i}$ phase shift and $\delta_i = \frac{A_{\text{obs}_i}}{A_{\text{th}_i}}$ the gravimetric factor (amplitude factor). The relation between the gravimetric factor and the Love numbers h_ℓ (for the displacement due to elastic deformation) and k_ℓ (for the potential of the deformed mass) is given by

$$\delta_\ell = 1 + \frac{2}{\ell} h_\ell - \frac{\ell + 1}{\ell} k_\ell. \tag{16}$$

For $\ell = 2$ it follows $\delta_2 = 1.16$ with $h_\ell = 0.61$ and $k_\ell = 0.3$. In Dehant (1987) one can find the Wahr–Dehant (WD) model tidal parameters for an elliptically uniformly rotating Earth with elastic inner core, a liquid inner core and an inelastic mantle. It is still in use for many applications. The most recent tidal models are the DDW model (Dehant et al. 1999) and the MAT01 (Mathews 2001; Mathews et al. 2002). Ducarme et al. (2002) compared these models and found the discrepancies between the models to be of the order of 0.1%.

The tidal deformation on the Earth's surface ξ at point P can be roughly expressed by

$$\xi = \frac{R \, h_2}{2 \, \delta_2 \, g} \delta g_{\text{ET}}, \tag{17}$$

with R = radius of the Earth, g = gravitational acceleration and δ_{ET} the gravity change due to the Earth tides at point P (Hinderer and Legros 1989; Hinderer et al. 1991).

The program PREDICT (Wenzel 1996) calculates the tidal acceleration and the theoretical Earth tides based on model tidal parameters (δ and κ) for different tidal potential catalogues (e.g. HW95 and Tamura).

Besides the time dependency, the Earth tides are also latitude dependent. Figure 10.16 shows as example the diurnal wave M1, the semidiurnal wave M2 and the long-period wave M*f*. The sectorial semidiurnal waves have their maximum at the equator ($\phi = 0°$) whereas the maximum of the tesseral diurnal waves is at $\phi = 45°$. The long-period waves have a minimum at about $\phi = 34°$.

The Earth tides are composed of the different partial tidal waves. Figure 10.17 shows four long-period theoretical tidal waves SA, SSA, MM and MF for Potsdam site ($\phi = 52.3806°$N, $\lambda = 13.0682°$E, $h = 81$ m) and the resulting long-periodic tidal waves (LP) calculated with the program PREDICT from the ETERNA package (Wenzel 1996) using the Tamura tidal catalogue (Tamura 1987).

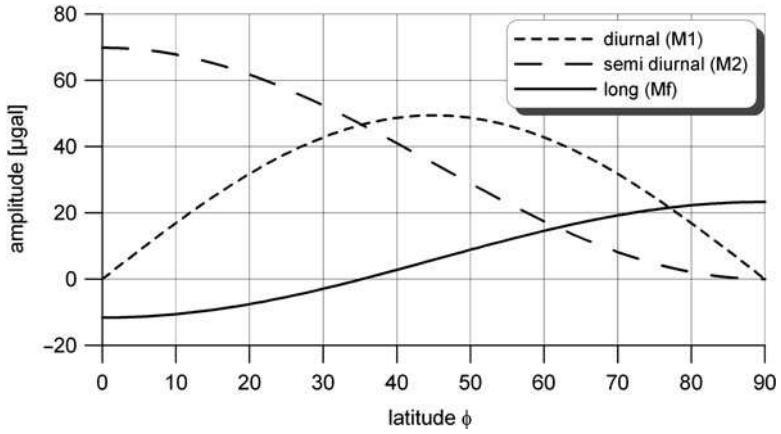


Fig. 10.16 Latitude dependency of the tidal waves

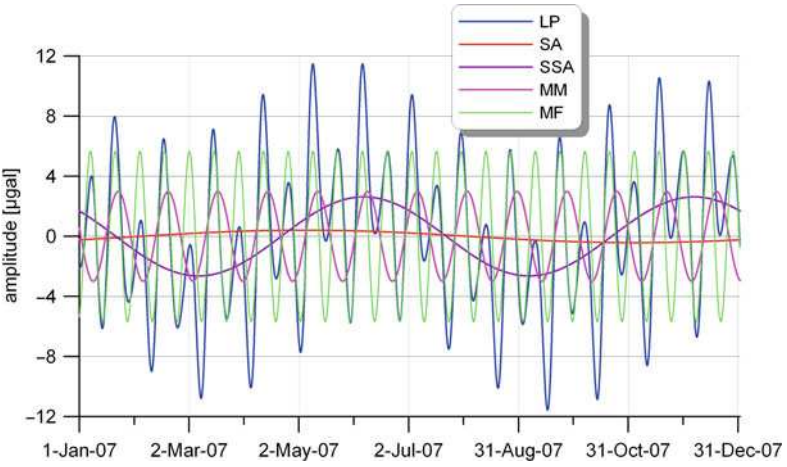


Fig. 10.17 Long-periodic model tides

The Earth tides are the largest signal in the recorded gravity data. They can reach up to ~280 µGal. Figure 10.14b shows an example for the ME station.

10.6.2 Gravity Variations Induced by the Atmosphere

The continual redistribution of air masses in the Earth's atmosphere causes temporal gravity variations. These variations highly correlate with the atmospheric pressure variations that can be measured and used for calculating this gravity effect. The period range of the ground pressure changes varies from minutes to years and may reach up to 60 hPa, which causes about 20 µGal gravity variations. This means the

magnitude of the air-masses-induced gravity variations can reach more than 10% of the tidal gravity effect. The redistribution of air masses affects the gravimeter sensor in two different ways:

1. Gravity changes due to the direct Newtonian attraction of the air masses (attraction term), rough estimation –0.43 μGal/hPa
2. Gravity changes due to vertical displacement of the gravimeter on the deformed Earth and the redistribution of masses caused by deformation (elastic deformation term), rough estimation 0.13 μGal/hPa

Both effects sum up to –0.3 μGal/hPa (Torge 1989). In the past, different methods for modelling the atmospheric pressure effect were developed which generally fall into two categories:

1. Empirical approaches

   These methods use the local atmospheric pressure for determining the single and complex admittance between atmospheric pressure and gravity data based on regression and cross-spectral methods (Warburton and Goodkind 1977; Crossley et al. 1995; Neumeyer 1995).
2. Physical approaches

   (a) 2D model
   These methods use global atmospheric pressure and temperature data measured at the Earth's surface and a standard altitude-dependent air density distribution. The physical approaches (Merriam 1992; Sun 1995; Kroner 1997; Boy et al. 1998; Kroner and Jentzsch 1998; Neumeyer et al. 1998; Vauterin 1998; Boy et al. 2002) based on atmospheric models determine the attraction and deformation terms according to Green's function (Farrell 1972).

   (b) 3D model
   These models use 3D global atmospheric pressure, humidity and temperature data for modelling of the attraction term (Neumeyer et al. 2004, 2007; Boy and Chao 2005; Klügel and Wziontek 2009). The deformation term is calculated according to the loading Green's function.

10.6.2.1 Empirical Methods

Because of the good correlation between the recorded local atmospheric pressure δap and the gravity δg variations, empirical mathematical methods can be used for the determination of the pressure–gravity admittance.

Single Admittance

The simplest method for the determination of the atmospheric pressure–gravity admittance adm_{ap} involves a linear regression between the atmospheric pressure

and the gravity variations. The atmospheric pressure-induced gravity variations δg_{air} can be calculated by $\delta g_{\text{air}} = \delta ap \times \text{adm}_{\text{ap}}$. The reduction of the atmospheric pressure effect with the single admittance coefficient delivers adequate results for many applications. However, it captures about 90% of the total effect only (Sun 1995). The goal is to capture the complete effect by applying other methods.

Figure 10.14c shows an example of the atmospheric pressure-induced gravity δg_{air} for the ME station calculated with an admittance coefficient of –0.313 μGal/hPa determined by means of linear regression.

Complex Admittance

This method determines the frequency-dependent atmospheric pressure–gravity admittance function cadm(f_n) in amplitude and phase. It is based on the cross-spectral analysis. For a single input–single output model the transfer function $\hat{H}(f)$ as a function of frequency can be calculated after Bendat and Piersol (1986) by

$$\hat{H}(f_n) = \frac{\hat{G}xy(f_n)}{\hat{G}xx(f_n)} \tag{18}$$

with the autospectral density $\hat{G}xx$ and the cross-spectral density $\hat{G}xy$ of the input series x_n and the output series y_n, sampled at equally spaced time intervals Δt. For evaluation of $\hat{G}xx$ and $\hat{G}xy$, the data records x_n and y_n must be divided into nb blocks each consisting of N data values. The block size N should be a power of 2, and then one can use the faster Cooley–Tukey procedure of fast Fourier transform (*fft*). If needed to suppress side-lobe leakage, the data values in each block can be tapered by applying an appropriate window, e.g. the Hanning window.

The averaged autospectral density estimates from the nb blocks of data can be calculated by the fast Fourier transform according to

$$\hat{G}xx_k = \frac{2}{nb\,N} \sum_i \left[\mathit{fft}(x_i)_n\right]^2 \tag{19}$$

with $i = 0,1,\ldots,nb–1$ and $k = 0,1,\ldots N/2$. The frequency is calculated by $f_k = k/(\Delta t \times N)$.

The averaged raw cross-spectral density estimates from the nb blocks of data can be calculated by

$$\hat{G}xy_k = \frac{2}{nb\,N} \sum_i \left[\mathit{fft}(x_i)^*_n\, \mathit{fft}(y_i)_n\right]. \tag{20}$$

The first term (*) is the complex conjugate of the Fourier transform of the record x_n and the second term the Fourier transform of the record y_n.

In this application the record x_i corresponds to the measured atmospheric pressure variations δap. The record y_i corresponds to $\delta g_{\text{SG}} = \delta g_{\text{pre}} - ET$ ($\delta g_{\text{pre}}=$

pre-processed gravity data, ET = theoretical Earth tides calculated with the local or WD model Earth tide parameters).

From the auto- and cross-spectral density the complex frequency admittance function cadm(f_n) can be calculated. The complex auto $\hat{G}xx(f_k)$ and cross-spectral density $\hat{G}xy(f_k)$ can be split into their real and imaginary parts and used for calculation of the gain cadm$A(f_n)$ and the phase cadm$\phi(f_n)$:

$$\mathrm{cadm}A(f_k) = \frac{\left\{\left[\mathrm{Re}(Gxy(f_k))\right]^2 + \left[\mathrm{Im}(Gxy(f_k))\right]^2\right\}^{1/2}}{\hat{G}xx(f_k)}, \tag{21}$$

$$\mathrm{cadm}\phi(f_k) = \tan^{-1}\left(\frac{\mathrm{Im}(Gxy(f_k))}{\mathrm{Re}(Gxy(f_k))}\right). \tag{22}$$

The admittance function cadm(f_k) is shown in Fig. 10.18. After determination of the frequency response function, the frequency-dependent atmospheric pressure correction can be carried out for the gravity data. This is possible in the frequency domain using *fft* techniques. The *fft* transforms the gravity data δg_{SG} and the atmospheric pressure data δap into the frequency domain, where the correction is done on the gravity data according to the complex admittance function. After that the inverse fast Fourier transform *ifft* retransforms the corrected gravity data δg_corr_ap back into time domain:

$$\delta g\_corr\_ap = ifft\left[fft(\delta g_{\mathrm{res}}) - fft(\delta ap)\ \mathrm{cadm}\right]. \tag{23}$$

The complex admittance gives improved correction of the atmospheric pressure effect over the total tidal band compared to the single admittance coefficient.

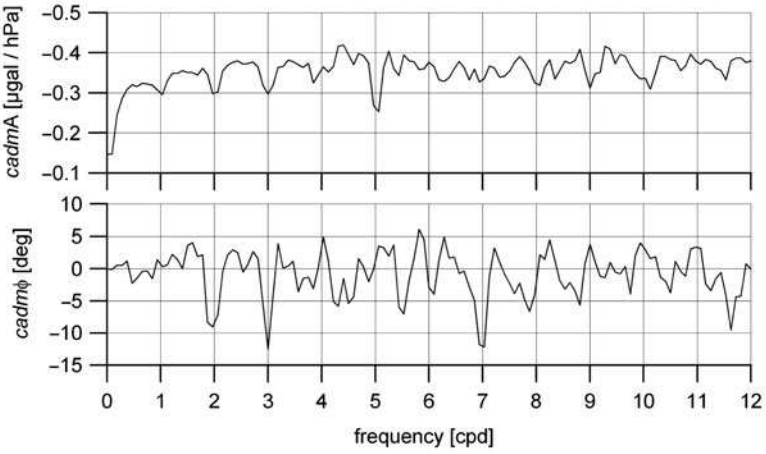


Fig. 10.18 Admittance function for gravity-atmospheric pressure (cadm(f))

10.6.2.2 Physical Models

Physical models are based on the Newtonian attraction of the air masses on the proof mass of the gravimeter and the Earth's elastic deformation due to the atmospheric loading. The accurate modelling of the attraction and deformation terms requires the knowledge of

- Air density distribution in space and time
- Atmospheric loading distribution in combination with a suitable Earth model

2D Model

This model is based on the calculation of gravity changes caused by a column load on the Earth's surface using the appropriate Green's functions, describing the response of a given elastic Earth model to a point load on its surface (Farrell 1972). The atmosphere is subdivided into columns and the altitude-dependent air density is considered by the standard atmosphere for mid-latitudes. The Green's functions for the Newtonian attraction GN(ψ) and deformation GE(ψ) terms have been calculated and tabulated by Merriam (1992) and Sun (1995) with ψ as the angular distance between the footprint centre of the column load on the Earth's surface and the gravimeter site.

If we inspect the Green's function (Fig. 10.19) for the attraction term, we notice a pronounced change of GN(ψ) in the range from $\psi = 0$ to about $\psi = 0.5°$ (local zone). Normally the atmospheric pressure data have a spacing of 0.5° and therefore the data must be interpolated.

In the local zone the attraction term is dominant. For a distance of $\psi = 1.1°$ the attraction and deformation terms are equal and there is no effect. For angular distances ψ >1.1° the deformation term becomes dominant and the attraction term is

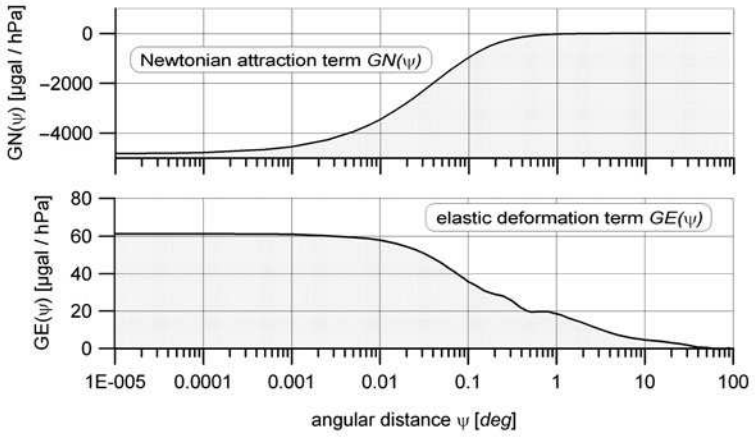


Fig. 10.19 Green's functions for attraction and deformation term (Sun 1995)

zero near $\psi = 2.7°$. For ψ >40° the deformation effect is small and can be neglected depending on the research goal.

The variations of the attraction and deformation term δg_{air} can be calculated from the Green's function according to (Sun 1995)

$$\delta g_{\text{air}}(t, \psi) = \left[\frac{\text{GN}(\psi) + \text{GE}(\psi)}{10^5 \psi} \frac{dS}{2\pi\,[1 - \cos(1°)]} \right] \delta ap(t, \psi, dS) \qquad (24)$$

with GN(ψ) and GE(ψ) = tabulated Green's functions for the Newtonian attraction and elastic deformation terms, ψ = angular distance from gravimeter to the air column in radian, dS = footprint of the air column and the surface atmospheric pressure variation δap at distance ψ for the surface area dS in radian^2.

The disadvantage of this method consists in using the standard altitude-dependent air density distribution. It does not consider the real air density variations.

3D Model

The 3D model considers the real altitude-dependent mass (density) distribution in the atmosphere for calculation of the attraction term. The deformation term can be calculated with 2D atmospheric pressure data according to the load Green's function for elastic deformation (cf. (24)).

From European Centre for Medium-Range Weather Forecasts (ECMWF) global 3D atmospheric pressure, humidity and temperature data are available with a spacing of 0.5° up to 60 km height and an interval of 6 h.

For determining the air masses, the air density ρ must be known. From the hydrostatic equations of dry air and water vapour (Etling 2002), (25) can be derived for calculating the air density ρ:

$$\rho = \frac{p}{R_L\, T \left\langle 1 - q + \frac{q}{\varepsilon} \right\rangle}. \qquad (25)$$

Input data for calculation of ρ are the ECMWF data for atmospheric pressure p, specific humidity q and air temperature T. The gas constant for dry air is R_L = 287 J/kg/K, and $\varepsilon = 0.622$ is the ratio of R_L and R_{W} (gas constant for water vapour). With this air density the attraction term can be calculated using an appropriate model.

Point Mass Model

This model (Neumeyer et al. 2004, 2007), based on the law of gravitation, calculates the gravitational acceleration at the gravimeter proof mass caused by mass redistribution within the atmosphere. For calculation the atmosphere is divided into spherical air segments sV. These air segments are approximated by point masses at segment centre positions (see Fig. 10.20).

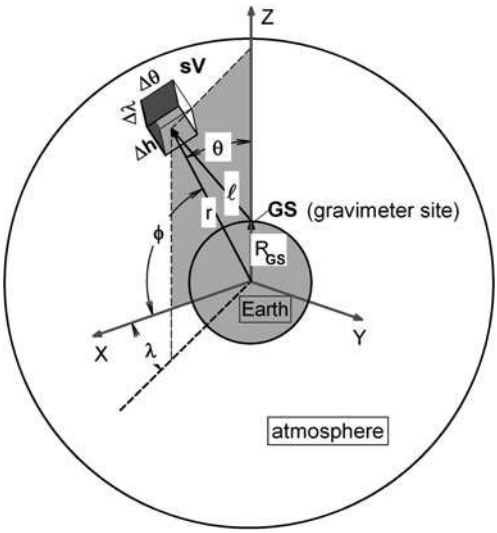


Fig. 10.20 Spherical air segment for calculating the air masses attraction; $r =$ geocentric distance, $\vartheta = (90 - \varphi) =$ polar distance, $\varphi =$ geocentric latitude, $\lambda =$ geographical longitude, GS = gravimeter site, $sV =$ air volume segment

The air density ρ of the air segments is calculated from ECMWF atmospheric pressure, humidity and temperature data (25). These data are dedicated to height h, latitude φ and longitude λ with the running indices ℓ, m, n. Equation (26) describes the vertical component of the gravitational acceleration δg_{air} which acts on the proof mass of the gravimeter caused by mass redistributions within the atmosphere:

$$\delta g_{air_{\ell,m,n}} = -G\left[\sum_{\ell,m,n} \rho_{\ell,m,n}\, sV_{\ell,m,n}\, A_{\ell,m,n}\right]$$

$$A_{\ell,m,n} = \frac{R_{\text{GS}} - r_\ell\,[\sin\vartheta_{\text{GS}}\,\sin\vartheta_m(\cos\lambda_{\text{GS}}\,\cos\lambda_n + \sin\lambda_{\text{GS}}\,\sin\lambda_n) + \cos\vartheta_{\text{GS}}\,\cos\vartheta_m]}{\left\{R_{\text{GS}}^2 + r_\ell^2 - 2R_{\text{GS}}r_\ell\,[\sin\vartheta_{\text{GS}}\,\sin\vartheta_m(\cos\lambda_{\text{GS}}\,\cos\lambda_n + \sin\lambda_{\text{GS}}\,\sin\lambda_n) + \cos\vartheta_{\text{GS}}\,\cos\vartheta_m]\right\}^{\frac{3}{2}}}, \tag{26}$$

where $G =$ gravitational constant, $\rho =$ density of the air segment, $sV =$ spherical air volume segment, $R_{\text{GS}} =$ Earth radius at the gravimeter station, $r_\ell = R_{\text{GS}} + h_\ell$, $\theta_{\text{GS}} =$ co-latitude of the gravimeter site, $\lambda_{\text{GS}} =$ longitude of the gravimeter site, θ and $\lambda =$ coordinates of sV.

Potential Model

This attraction model has been derived from the gravitational potential of the air masses. For calculation the atmosphere is divided into spherical air segments sV and for mathematical reasons (transformation of the spherical into a Cartesian coordinate system) a coordinate system is introduced in such a way that the north pole of this system coincides with the gravimeter site (GS). The gravitational potential of the air masses is calculated with respect to the gravimeter site. By partial derivation of the gravitational potential of the air masses and summation of overall

spherical air segments we obtain the vertical component of the gravitational acceleration $\delta g_{\mathrm{air}}(R_{\mathrm{GS}}, 0, 0)$ (27) caused by mass redistributions within the atmosphere.

Because of the wide spacing (0.5°) the ECMWF data must be interpolated, which is accomplished by a two-dimensional (in θ and λ) bi-linear interpolation. The amount of interpolated points of 10×10 can be regarded as adequate. More details can be found in Neumeyer et al. (2004, 2007):

$$\delta g_{air}(R_{\mathrm{GS}}, 0, 0) = -G \sum_{\ell,m,n} \rho_{\ell,m,n}(\lambda_n - \lambda_{n-1}) \frac{1}{6R_{\mathrm{GS}}^2} B_{\ell,m}, \tag{27}$$

$$B_{\ell,m} = \begin{bmatrix} \sqrt{r_{\ell-1}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell-1}R_{\mathrm{GS}}\cos\vartheta_{m-1}} \cdot \left(2r_{\ell-1}^2 - R_{\mathrm{GS}}^2 + 2R_{\mathrm{GS}}r_{\ell-1}\cos\vartheta_{m-1} + 3R_{\mathrm{GS}}^2\cos 2\vartheta_{m-1}\right) \\ -\sqrt{r_{\ell}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell}R_{\mathrm{GS}}\cos\vartheta_{m-1}} \cdot \left(2r_{\ell}^2 - R_{\mathrm{GS}}^2 + 2R_{\mathrm{GS}}r_{\ell}\cos\vartheta_{m-1} + 3R_{\mathrm{GS}}^2\cos 2\vartheta_{m-1}\right) \\ -\sqrt{r_{\ell-1}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell-1}R_{\mathrm{GS}}\cos\vartheta_{m}} \cdot \left(2r_{\ell-1}^2 - R_{\mathrm{GS}}^2 + 2R_{\mathrm{GS}}r_{\ell-1}\cos\vartheta_{m} + 3R_{\mathrm{GS}}^2\cos 2\vartheta_{m}\right) \\ +\sqrt{r_{\ell}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell}R_{\mathrm{GS}}\cos\vartheta_{m}} \cdot \left(2r_{\ell}^2 - R_{\mathrm{GS}}^2 + 2R_{\mathrm{GS}}r_{\ell}\cos\vartheta_{m} + 3R_{\mathrm{GS}}^2\cos 2\vartheta_{m}\right) \\ -1n\left[r_{\ell-1} - R_{\mathrm{GS}}\cos\vartheta_{m-1} + \sqrt{r_{\ell-1}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell-1}R_{\mathrm{GS}}\cos\vartheta_{m-1}}\right] 6R_{\mathrm{GS}}^3\cos\vartheta_{m-1}\sin^2\vartheta_{m-1} \\ +1n\left[r_{\ell} - R_{\mathrm{GS}}\cos\vartheta_{m-1} + \sqrt{r_{\ell}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell}R_{\mathrm{GS}}\cos\vartheta_{m-1}}\right] 6R_{\mathrm{GS}}^3\cos\vartheta_{m-1}\sin^2\vartheta_{m-1} \\ +1n\left[r_{\ell-1} - R_{\mathrm{GS}}\cos\vartheta_{m} + \sqrt{r_{\ell-1}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell-1}R_{\mathrm{GS}}\cos\vartheta_{m}}\right] 6R_{\mathrm{GS}}^3\cos\vartheta_{m}\sin^2\vartheta_{m} \\ -1n\left[r_{\ell} - R_{\mathrm{GS}}\cos\vartheta_{m} + \sqrt{r_{\ell}^2 + R_{\mathrm{GS}}^2 - 2r_{\ell}R_{\mathrm{GS}}\cos\vartheta_{m}}\right] 6R_{\mathrm{GS}}^3\cos\vartheta_{m}\sin^2\vartheta_{m} \end{bmatrix},$$

where G = gravitational constant, ρ = air density as function of the spherical coordinates (r, θ, λ), R_{GS} = Earth radius at gravimeter's site.

Results of 3D Attraction Models

By applying the point mass and potential model it could be shown that point and potential model deliver the same results within small error bars of about ±0.1%.

The application of the 3D attraction model indicates a seasonal surface pressure-independent (SPI) gravity effect caused by movements of air masses without changes of the surface atmospheric pressure (Meurers 1999; Simon 2002). In the summer season the air masses move up and the atmospheric attraction term decreases, whereas in winter the air masses move down, causing a larger attraction term. Figure 10.21 illustrates this.

The gravity effect caused by mass relocation is different above and to the side of the gravimeter. When we assume a mass relocation from point 1 to point 2 above the gravimeter, the attraction term becomes smaller ($g_{\mathrm{air}}\_2 < g_{\mathrm{air}}\_1$). When we assume a mass relocation from point 1a to point 2a to the side of the gravimeter the vertical component of the attraction term, which acts on the gravimeter proof mass, becomes larger ($g_{air}\_2\mathrm{a} > g_{air}\_1\mathrm{a}$). This effect is altitude dependent.

By applying the 3D attraction model (Neumeyer et al. 2004, 2007), it could be shown that the gravity variations induced by mass redistribution in the atmosphere include the SPI gravity effect in the order of 1–2 μGal. Figure 10.22 shows an example of the SPI effect at the same ground pressure of 1001.25 ± 0.25 hPa calculated for the coordinates of the Metsahovi (ME) SG station.

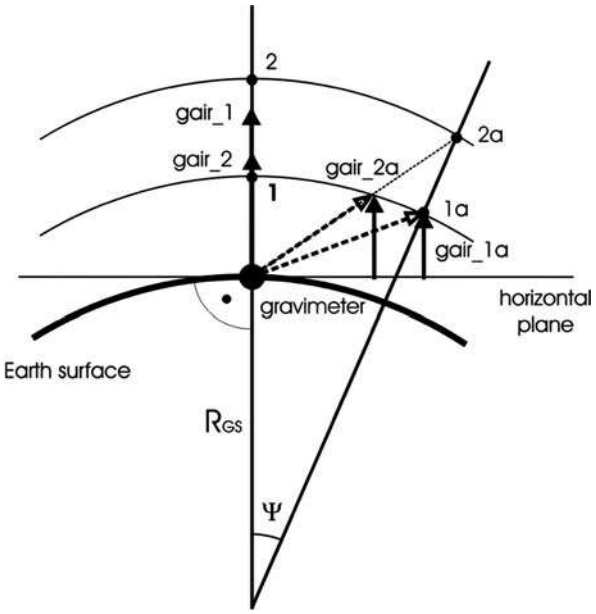


Fig. 10.21 Mass relocation within the atmosphere

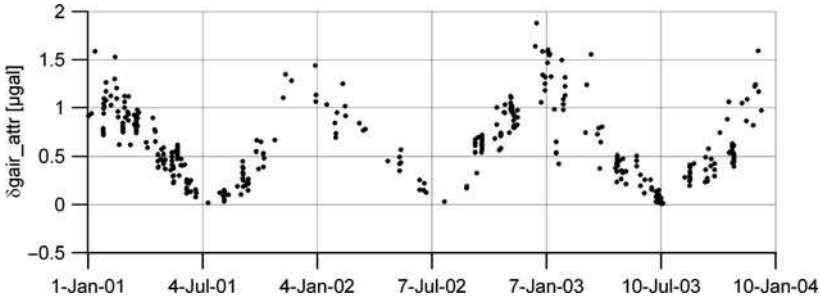


Fig. 10.22 Surface pressure-independent gravity effect (SPI) at surface pressure of 1,001.25 ± 0.25 hPa at SG site ME

Recent studies (Klügel and Wziontek 2009) showed that the seasonal SPI effect becomes smaller or disappears when using atmospheric data up to 84 km height. The accuracy of the calculation of the attraction and deformation term strongly depends on the accuracy of the atmospheric model (e.g. ECMWF).

10.6.3 Hydrology-Induced Gravity Variation

Any redistribution of water masses induces changes of the Earth's gravity field. Precipitation (rain and snow) causes mainly changes in soil moisture and groundwater level and also in surface water, e.g. in Amazon area. One part of this precipitation evaporates at the surface of the Earth and in the ground. Another part increases the

groundwater level or drains off. The water circulation in the surroundings of the gravimeter causes gravity variation by gravitational attraction on the proof mass of the gravimeter and the deformation of the Earth's surface. Both terms result in gravity variation similar to that due to the atmosphere. The attraction term reflects the local and the global hydrological changes. The deformation term is mostly caused by global hydrological changes.

For calculation of the hydrology-induced gravity variations, different models are in use. These models can be divided into empirical and physical approaches.

The empirical methods can generally be divided into Bouguer plate modelling and the determination of an admittance coefficient between gravity residuals and groundwater level variations or soil moisture, etc.

The physical methods use the local and global water mass redistributions from local measurements or local and global hydrological models. They calculate the Newtonian attraction of the water mass changes on the proof mass of the gravimeter (attraction term) and the deformation of the Earth caused by the changes of the water masses at global scale (deformation term). The most challenging problem is obtaining usable data from the water mass changes. The changes of the water masses around the SG at local or global scale are measured by groundwater level (gwl) and soil moisture (sm) changes as well as precipitation. These measurements and additional meteorological, terrain and ground parameters are input for the local and global hydrological models which deliver the water mass changes for the local or global area in an adequate gridding. Up to now mainly global hydrological models are available.

10.6.3.1 Bouguer Plate Model

A first estimation of δg_{gwl} can be calculated by the Bouguer plate modelling (Torge 1989):

$$\delta g_{\mathrm{gwl}} = 2\,\pi\; G\; \rho_{\mathrm{w}}\; \mathrm{Ps}\; \delta h\_gwl \tag{28}$$

with the gravitational constant $G = 6.673 \times 10^{-11}\mathrm{m}^3/\mathrm{kg/s}^2$, the water density $\rho_{\mathrm{W}} = 1\ \mathrm{gcm}^3$, the water-filled pore space Ps and the groundwater level changes δh_gwl. If we assume $\delta h\_gwl = 1$ m and Ps = 10%, δg_{gw} reaches 4.2 μGal. This corresponds to an admittance of $\mathrm{adm}_{\mathrm{gwl}} = 4.2$ μGal/m. With the same equation the soil moisture-induced gravity effect can be estimated when we replace δh_gwl by the thickness of the moist layer and Ps by the percentile soil moisture change. The problem of this method consists in the very imprecise estimation of the water-filled pore space.

10.6.3.2 Precipitation Model

Groundwater level accumulation caused by precipitation as rainfall δh_rain can be modelled according to (29) and (30) (Crossley and Su 1998):

$$\delta h\_r(t, t1) = \mathrm{rain}(t1)\,(1 - \mathrm{e}^{(t1-t)/\tau 1})\,\mathrm{e}^{(t1-t)/\tau 2} \tag{29}$$

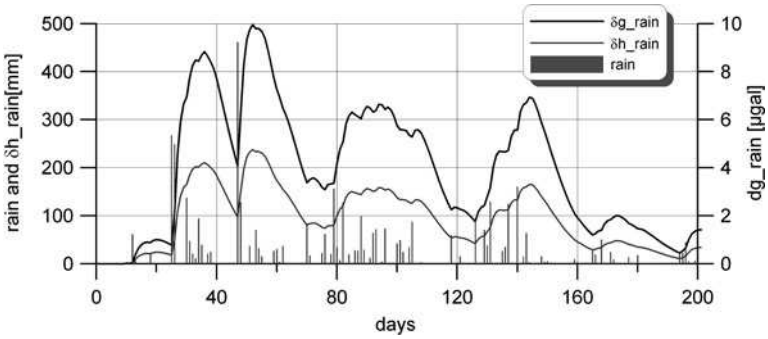


Fig. 10.23 Modelling of precipitation

with rain = amount of rain at time $t1 = 0...tn$, time $t = 0...tn$, $\tau 1$ = recharging time constant, $\tau 2$ = discharging time constant and the additional condition $\delta h\_r(t,t1) = 0$ if $\delta h\_r(t,t1)<0$ which yields

$$\delta h\_rain(t) = \sum_{t1} \delta h\_r(t, t1). \tag{30}$$

The gravity caused can be calculated according to (28) by the Bouguer plate model. Figure 10.23 shows an example where the accumulated groundwater level δh_rain and δg_rain are calculated from rainfall data for $\tau 1 = 6$ h and $\tau 2 = 8$ days. The time constants $\tau 1$ and $\tau 2$ must be adapted empirically depending on the geological structure of the site and its surrounding. This model yielded satisfactory results at some SG stations (Crossley and Su 1998; van Dam and Francis 1998; Lambert et al. 2006).

10.6.3.3 Single Admittance Model

Groundwater level changes δh_gwl, normally the main part of the water masses change, soil moisture changes δsm and precipitation correlate with the SG-measured gravity variation δg_{SG}. Therefore the groundwater level admittance on gravity $\mathrm{adm}_{\mathrm{gwl}}$ can be determined by a linear regression analysis and the groundwater level-induced gravity variation δg_{gwl} can be calculated by $\delta g_{\mathrm{gwl}} = \delta h\_gwl \times \mathrm{adm}_{\mathrm{gwl}}$. The same calculation can be carried out for soil moisture.

Presently, most of the SG stations are equipped with a borehole for measuring groundwater level variations. In many cases a good correlation between gravity and groundwater level variations could be shown (Kroner 2001; Harnisch and Harnisch 2002; Virtanen 2001, 2006). The gravity effect of these variations was determined by the regression analysis. The admittance coefficient can reach several microgalileos per metre depending on the hydrological conditions. This is a simple model, which does not reflect the real hydrological gravity signal very accurately.

While suitable approaches exist for atmospheric signals (cf. Sect. 10.6.2.2), the problem is much more difficult and not yet satisfactorily solved when regarding hydrological signals. Both soil moisture and groundwater level data reflect local effects as well as signals on regional or continental scale. Additionally, topography and local hydrological structure have a big impact on hydrological gravity effects (e.g. Boy et al. 2005; Virtanen et al. 2005; Kroner and Jahr 2006; Meurers et al. 2005, 2007).

Figure 10.14e shows the hydrological gravity effect based on groundwater level variations near the SG station calculated with the admittance coefficient of 3.3 μGal/m.

10.6.3.4 Global Hydrological Models

Global hydrological models can be used to derive the associated gravity variation for a SG site at global scale. The global hydrological models try to represent both the spatial distribution and the changes of continental water budget with time. The water budget change (wbc) of all global hydrological models is based on precipitation (prec), runoff (roff) and evapotranspiration (evpt) expressed by the very simple fundamental relation:

$$\text{wbc} = \text{prec} - \text{roff} - \text{evpt}. \tag{31}$$

However, the implementation brings difficulties, since the water stocks cannot be measured directly. As a consequence different hydrological models could be quite different, especially from the point of view of reflecting entirely the redistribution of water masses. Several hydrological models are available, e.g. Water Gap Global Hydrology Model (WGHM) (Döll et al. 2003), Leaky-Bucket Model (H96) (Huang et al. 1996; Fan and Van den Dool 2004), Land Dynamics Model (LaD) (Milly and Shmakin 2002).

The output of these three models is available in the form of gridded data sets representing monthly averages of water storage expressed as equivalent water columns in millimetre or centimetre. The grid step in geographical latitude and longitude is 0.5° for the first two models and 1° for the third one. The models are updated regularly so that the current data are available with a delay of several months. It should be stressed that the depth of the considered groundwater in different models differs; the snow is taken into account only partially and in different ways, and the modelling of ice shields is, according to the authors, very incomplete and unreliable.

The hydrologically induced gravity variations (global attraction and deformation terms) can be derived from the global hydrological models as follows. Changes in terrestrial water storage expressed in equivalent water thickness are expanded into spherical harmonics, transformed into potential coefficients (e.g. up to degree $\ell_{\max} = 10$) and then used for calculating the gravity changes at the SG location (Neumeyer et al. 2006).

Input data for each epoch are given on a regular geographical grid as equivalent water thickness $\delta h(\varphi, \lambda)$ which represents deviations from some reference state of

the water budget in the considered block. Since $\delta h(\varphi, \lambda)$ does not represent some real water column, but an equivalent of complete water mass excess or deficit contained in groundwater, surface water, soil moisture and other considered components, it is reasonable to regard it as surface mass using the relation

$$\delta\sigma(\varphi, \lambda) = \rho_{\mathrm{W}} \delta h(\varphi, \lambda), \tag{32}$$

where ρ_{W} is the density of fresh water.

In order to compute the gravitational effects, a numerical integration of the surface mass density $\delta\sigma$ is necessary. This can be done according to Wahr et al. (1998). Furthermore, the variation δr of the radial position of the SG station caused by the load must be considered. The equation for gravity variations derived from hydrological model δg_{HM} then takes the form

$$\delta g_{\mathrm{HM}}(\varphi, \lambda) = \frac{GM}{R^2} \sum_{\ell=0}^{\ell\,\mathrm{max}} (\ell + 1 - 2h'_\ell) \sum_{m=0}^{\ell} \left[\delta\overline{C}_{\ell m}^{\mathrm{HM}} \cos(m\lambda) + \delta\overline{S}_{\ell m}^{\mathrm{HM}} \sin(m\lambda) \right] \overline{P}_{\ell m}(\sin\varphi), \tag{33}$$

where φ, λ are the spherical geocentric coordinates of the computation point (longitude, latitude), R is the reference radius (mean equatorial radius of the Earth), GM is the gravitational constant times mass of the Earth, ℓ, m are degree and order of the spherical harmonics, $\ell_{\max}$ is chosen maximum degree in practical calculations (any natural number, $\ell_{\max} < \infty$), h'_ℓ are degree-dependent load Love numbers (Farrell 1972; Zürn and Wilhelm 1984; Hinderer and Legros 1989). $\overline{P}_{\ell m}$ are the fully normalized Legendre functions and $\overline{C}_{\ell m}, \overline{S}_{\ell m}$ are the fully normalized Stokes' coefficients. Superscript HM is related to spherical harmonic coefficients based on gravity variations derived from hydrological model. More details can be found in Neumeyer et al. (2006).

For arbitrary $\ell_{\max}$ degree-0 ($\ell = 0$) and degree-1 ($\ell = 1$) terms cannot be used in (33), since the hydrological models cannot and do not contain mass conservation, which does not exist on the level of continental water budget alone, and since hydrologically derived geocentre variations are not interpretable for the same reason.

Figure 10.24 shows the gravity variations δg_{HM} derived from different hydrology models calculated according to (33) up to $\ell_{\max} = 10$ at the SG station Bad Homburg (BH), Germany (50.2285°N, 8.6113°E and 190 m). The differences among the three models show the difficulties in modelling the gravity effect based on global hydrological models and point to their uncertainties. However, they give an estimation of the global hydrological signal, which can be interpreted as the global attraction and deformation term of the hydrological effect (Boy and Hinderer 2006). The attraction term is underestimated because of the smoothing of the local hydrological data (spherical harmonic expansion up to some low degree ($\ell_{\max} = 10$).The deformation term is well considered because of using the global hydrological loading (global hydrological data).

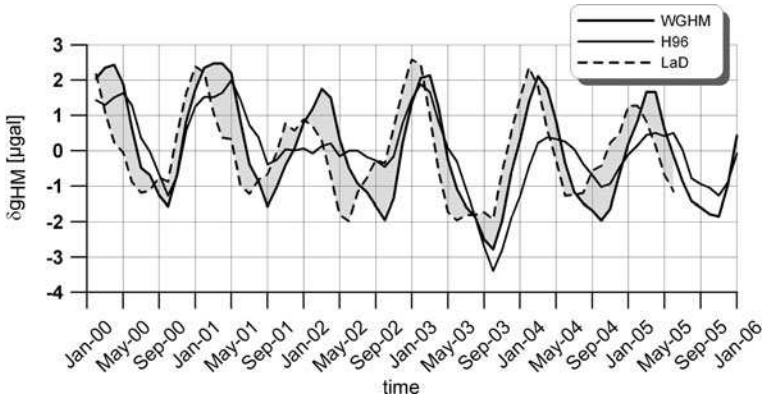


Fig. 10.24 Gravity variations δg_{HM} derived from hydrology models WGHM, H96 and LaD at the SG site Bad Homburg, Germany

For comparison, the deformation effect, the hydrological loading, was calculated independently with the mass loading Green's function GE(ψ) (cf. Fig. 10.19) and the data from the global hydrological model (WGHM). These data are delivered as millimetre water column (1 mm water column corresponds to 1 kp/m^2 = 0.981 hPa). Equation (24) was used to calculate the deformation term due to the hydrological loading $\delta g_{\mathrm{HM\_load}}$ (Fig. 10.25). The deformation term is more than 50% of the total effect (attraction plus deformation term).

Global and, for some areas, regional hydrological models are presently the models for calculating the deformation term, whereas the attraction term can be modelled according to the 3D model (e.g. point mass model) for the atmosphere

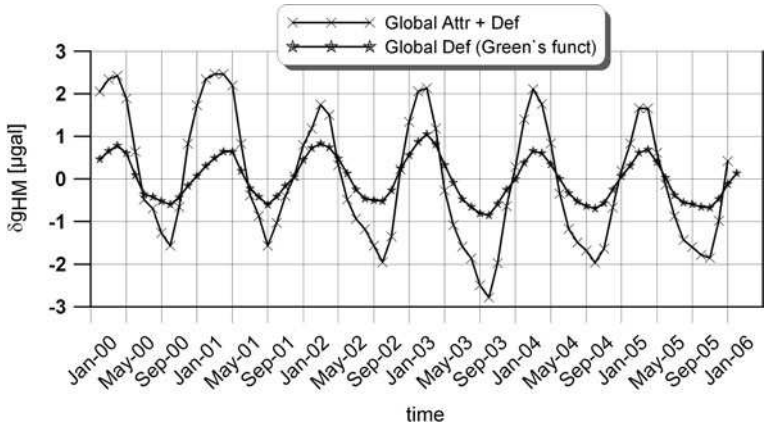


Fig. 10.25 Gravity variations derived from hydrology model WGHM, at the SG site Bad Homburg, Germany. δg_{HM} = global attraction and deformation term, $\delta g_{\mathrm{HM\_load}}$ = global deformation term

(cf. section "Potential Model"). Unfortunately, usually less data are available. To obtain the data a local hydrology model is required around the SG site, which considers the local hydrological cycle (Creutzfeldt et al. 2008; Kroner and Jahr 2006).

10.6.4 Ocean Tide Loading Gravity Effect

The ocean contributes to the noise at the site and causes the ocean loading effect, an elastic deformation of the Earth, combined with a gravity effect at the SG site. By low-pass filtering the ocean noise can be reduced considerably. The loading effect caused by the ocean tides superimposes the gravity signal as perturbation. Therefore, it is necessary to remove it as completely as possible from the gravity data before they are used for the study of geophysical and geodynamical effects.

The modelling of the ocean tide loading is based on global ocean tide models like FES2002 (Lefevre et al. 2002; Le Provost et al. 2002), NAO99b (Matsumoto et al. 2000, 2001) and the Green's function. The calculation of the ocean tides gravity effect at the SG site can be carried out according to

$$\mathrm{ol}_j(\varphi,\lambda) = \rho_W \sum_{i=0}^{N} \left(G(\psi_i)\overline{Z}_{i,j}(\varphi',\lambda')dS_i \right), \tag{34}$$

where ol_j is the gravity effect of the ocean tide wave j, e.g. M2 at the SG site with the coordinates φ and λ, ρ_W is the mean density of seawater, $G(\psi)$ Green's function for gravity acceleration or radial displacement, ψ angular distance between the ocean point mass load with the coordinates φ' and $\lambda'$and the SG site with the coordinates φ and λ, $\overline{Z}(\varphi',\lambda') = Ze^{i\mathrm{ph}}$ is the complex form of the ocean tide (amplitude Z and phase ph) and dS is the surface area of the mass load.

For the numerical evaluation of (34) the ocean is divided into a set of cells with assigned ocean tide amplitude and phase $\overline{Z}(\varphi',\lambda') = Ze^{i\mathrm{ph}}$ for each cell, which corresponds to a column load at the surface area dS. This information is provided by the ocean tide models as co-tidal maps for the main tidal waves with a gridding (distance) of the cells of $0.5° \times 0.5°$, and N is the total number of the oceanic cells.

The Green's functions represent the response of the Earth to a point load using an appropriate Earth model, e.g. PREM model. Figure 10.26 shows the Green's functions which are used for calculation of the ocean loading in the program LOAD 97 (Francis and Mazzega 1990). The Green's functions determine how much the Earth is deformed due to the point load of 1 kg as function of the angular distance to the SG station.

Figure 10.14f shows the calculated gravity variations induced by the ocean loading (δg_{ol}) using data from the ocean tidal model FES2002.

More information and programs for calculation of the ocean loading can be found at http://www.oso.chalmers.se/~loading/index.html and http://www.miz.nao.ac.jp/staffs/nao99/index_En.html

Penna et al. (2008) give an assessment of the accuracy of ocean tide loading.

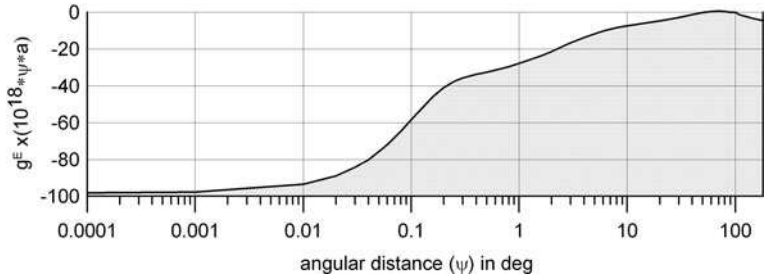


Fig. 10.26 Mass loading Green's function for the PREM Earth model (surface loading 1 kg)

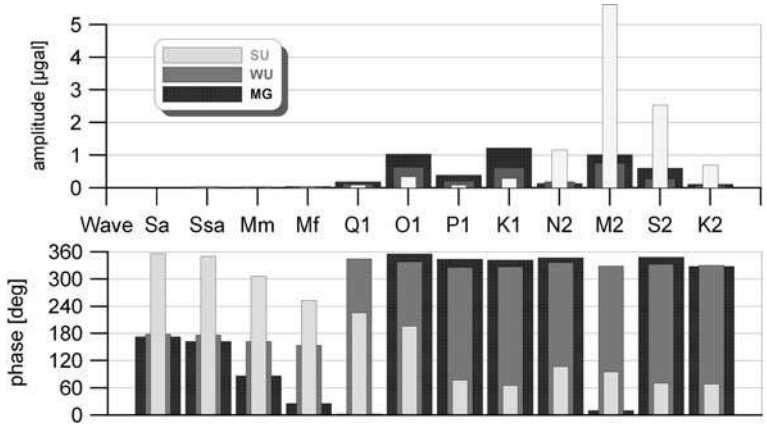


Fig. 10.27 Ocean tidal loading amplitudes and phases at SG stations in SU, WU and MG

10.6.4.1 Ocean Tide Loading Correction of the Tidal Parameters

The ocean tidal loading can be modelled and adjusted for the tidal parameters according to a suitable ocean tide model. Based on the global ocean tide models, e.g. FES2002 (Lefevre et al. 2002; Le Provost et al. 2002), NAO99b (Matsumoto et al. 2000, 2001), the ocean loading vectors for various tidal waves in diurnal-, semidiurnal- and long-periodic bands can be calculated with the program LOAD97 (Francis and Mazzega 1990). Figure 10.27 shows examples for the ocean tidal loading vectors for the SG stations in SU and Wuhan (WU) in China (30.5159°E, 114.4898°N and height 80 m) and MG.

The relations among the vectors (e.g. for the tidal wave M2), observed tide $\vec{A}\,(A,\,\alpha)$, theoretical tide $\vec{R}\,(\vec{R}, \alpha = 0)$, ocean tide loading $\vec{L}\,(L, \lambda)$, observed residuals $\vec{B}\,(B, \beta)$ and the remaining residuals $\vec{X}\,(X, \xi)$ after ocean loading correction, are given in Fig. 10.28.

After subtraction of the ocean tidal loading gravity vector $\vec{L}\,(L, \lambda)$ from the observed tidal vector $\vec{A}\,(A, \alpha)$, the remaining final residuals $\vec{X}\,(X, \xi)$ are still significant. Their amplitudes and phases can be calculated according to (35) and (36).

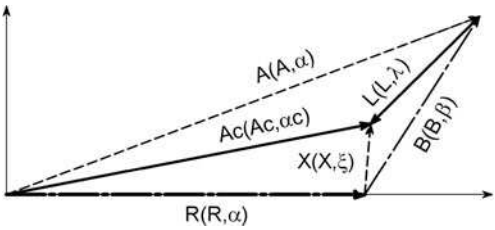


Fig. 10.28 Vector diagram of Earth tide and ocean loading vectors. $\vec{R}\left(\vec{R}, \alpha = 0\right) =$ theoretical tides, $\vec{A}\,(A, \alpha) =$ observed tides, $\vec{B}\,(B, \beta) =$ observed residuals, $\vec{L}\,(L, \lambda) =$ ocean loading, $\vec{A}_c\,(A_c, \alpha_c) =$ corrected tides, $\vec{X}\,(X, \xi) =$ corrected residuals

It can also be seen that the residuals $\vec{X}\,(X, \xi)$ are smaller than the observed residuals $\vec{B}\,(B, \beta)$ and the corrected observed tidal vector $\vec{A}_{\mathrm{c}}(A_{\mathrm{c}}, \alpha_{\mathrm{c}})$ is closer to the theoretical tides in amplitude and phase:

$$X_i = \left[B_i^2 + L_i^2 - 2B_i L_i \cos(\beta_i - \lambda_i)\right]^{1/2}, \tag{35}$$

$$\xi_i = \tan^{-1}\left[\frac{B_i \sin\,\beta_i - L_i \sin\,\lambda_i}{B_i \cos\,\beta_i - L_i \cos\,\lambda_i}\right]. \tag{36}$$

With the residual vector $\vec{X}\,(X, \xi)$, the WD model tidal amplitude R^{th} and the model δ^{th} factor, the ocean loading-corrected Earth tide parameters δ^{c} and κ^{c} can be calculated with (37) and (38) (Melchior et al. 1981; Sun et al. 1999):

$$\delta_i^{\mathrm{c}} = \delta_i^{\mathrm{th}}\left[1 + \left(\frac{X_i}{R_i^{\mathrm{th}}\delta_i^{\mathrm{th}}}\right)^2 + 2\left(\frac{X_i}{R_i^{\mathrm{th}}\delta_i^{\mathrm{th}}}\right)\cos\,\xi_i\right]^{1/2}, \tag{37}$$

$$\kappa_i^{\mathrm{c}} = tan^{-1}\left[\frac{X_i \sin\,\xi_i}{R_i^{\mathrm{th}}\delta_i^{\mathrm{th}}} + X_i \cos\,\xi_i\right]. \tag{38}$$

Several analyses (e.g. Sun et al. 1999; Ducarme et al. 2002; Neumeyer et al. 2005) demonstrate that the main reasons for the remaining residuals $\vec{X}\,(X, \xi)$ are the load vectors $\vec{L}\,(L, \lambda)$, which are not well determined due to the inaccuracy of the global ocean tide models, the lack of regional models and the complicated bay coastal lines which induce a kind of special shallow sea tidal phenomena (Francis 1992; Sun et al. 2002a). This means that the ocean models do not reflect well the real ocean tidal phenomena. The remaining residual vector $\vec{X}\,(X, \xi)$ after ocean loading correction can be of the order of some microgalileos for the main tidal waves depending on the SG location (e.g. 1.8 μGal for M2 at SG station in Sutherland). To improve the loading correction and reduce the discrepancy between observed and predicted ones, additional tide gauge measurements can be helpful. This is recommended for SG stations near the ocean with a strong ocean influence (Khan and Hoyer 2004; Sun et al. 2006).

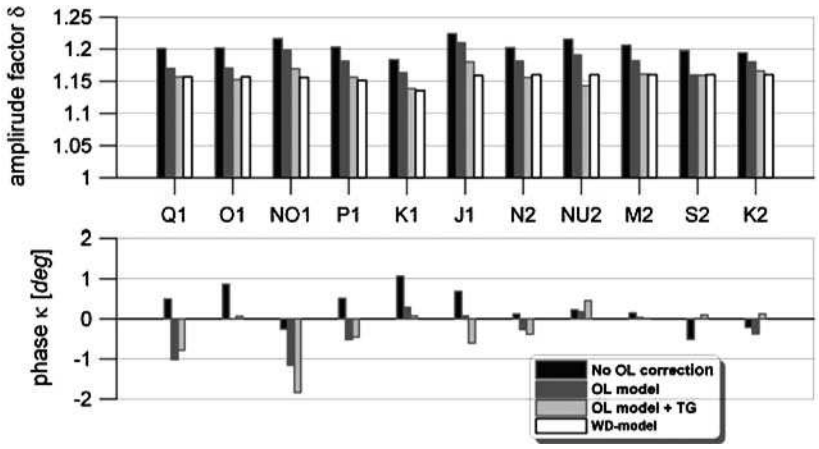


Fig. 10.29 Different ocean loading corrections of the tidal parameters in comparison with the WD model parameters. *White* = WD model, *black* = no ocean loading correction, *dark grey* = NAO99b and FES2002 (OL model), *light grey* = NAO99b and FES2002 and additional tide gauge ocean loading (OL model + TG), station Rio Carpintero, Cuba

In Neumeyer et al. (2005), analysis results are shown with additional tide gauge measurements. Precise tidal gravity (LaCoste and Romberg gravimeter) and atmospheric pressure observations were carried out at station Rio Carpintero, Cuba, in combination with tide gauge measurements at the coast of Santiago de Cuba. It was found that the ocean loading influence can reach about 3% in the diurnal band and 2% in the semidiurnal tidal band. Therefore, additional tide gauge measurements were carried out to improve the ocean loading correction. This improvement is based on the admittance (cf. section "Complex Admittance") between the tide gauge measurements and the gravity residuals corrected for the ocean loading based on the NAO99b and FES2002 models. Figure 10.29 shows as example different ocean loading corrections of the tidal parameters at station Rio Carpintero in comparison with the WD model parameters. Here the additional tide gauge corrected tidal parameters are much closer to the WD model parameters than those corrected with ocean tide models only. Because this is an empirical model only, a local ocean model should deliver better results.

10.6.5 Polar Motion

The main components of the polar motion are the free oscillation of the Chandler wobble with a period of about 435 days and an annual enforced oscillation mainly caused by seasonal mass redistributions in the atmosphere and the ocean (Gross et al. 2003).

When we assume the Earth to be a gyroscope, the polar motion can be explained according to Fig. 10.30 with the space-fixed angular momentum axis H, the figure

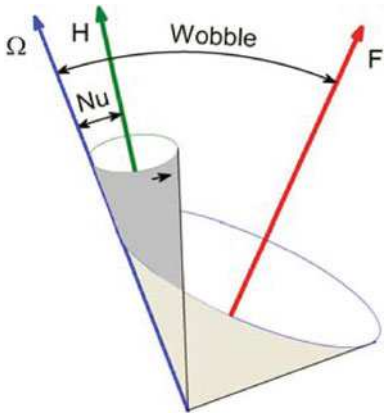


Fig. 10.30 Polar motion (Chandler wobble)

axis of the Earth F and the instantaneous Earth rotation axis Ω. The angular momentum axis is assumed as fixed. The instantaneous rotation axis Ω describes one cone around the figure axis F (body cone) and a second one (space cone) around the axis of angular momentum H. The motion can be described by a winding of the body cone around the space cone without slipping. The axes Ω, H and F are always coplanar. For an oblate gyroscope like the Earth the Ω axis and the figure axis F span a plane which revolves around the axis of angular momentum. The instantaneous rotation axis Ω is located where the two cones touch each other. The motion of the figure axis describes the Chandler wobble. The angles of the cones represent the amplitudes of wobble and nutation.

Whereas the Chandler wobble has a long period of 435 days, its nutation has a short period and small amplitude. The polar motion causes changes in centrifugal acceleration at the SG site, which can be measured with the SG.

The recording of the polar motion is carried out with VLBI, LLR, SLR and GPS measurements. The International Earth Rotation and Reference Systems Service (IERS) (http://www.iers.org) provides a smoothing of these measurements with a resolution of 1 day. From the IERS polar motion data $X_P(t)$ and $Y_P(t)$ (in terrestrial frame) the gravity effect of the polar motion can be calculated for the SG station with co-latitude θ and longitude λ according to (Torge 1989)

$$\delta g_{\mathrm{Pol}}(\theta, \lambda, t) = R\omega^2 \delta_{\mathrm{Pol}} \sin(2\theta)\ [X_P(t)\cos(\lambda) - Y_P(t)\sin(\lambda)] \qquad (39)$$

with R = radius of spherical Earth model, ω= angular velocity of the Earth according to the IERS Conventions 2003 (McCarthy and Petit 2004) and δ_{Pol}= 1.16 gravimetric factor of long periodic tides (Wahr 1985; Xu et al. 2004). The polar motion can reach a gravity effect of up to ~10 μGal which corresponds to about 1.2 arcs. For mid-latitudes $\theta = 45°$ dg_{Pol} has a maximum. An example of the polar motion calculated for IERS data is shown in Fig. 10.14d for the Metsahovi station.

10.6.6 Instrumental Drift

In spite of high persistent magnetic field levitating the proof mass (sphere) of the SG and the stabilized feedback system which brings the sphere in zero position, the SG has a drift of some microgalileos per year. The drift of the SG shows—after about 2 months of installation—an exponential behaviour. Shortly after this the drift curve becomes practically linear. As the main reasons for the drift, the magnetic field variations, gas adsorption on the levitating sphere or helium gas pressure variations around the sphere within the gravity sensing unit are assumed (Van Camp and Francis 2007). Because the drift biases long-term gravity variations, the drift must be determined and subtracted from the gravity signal.

As drift can be assumed, the remaining gravity signal cannot be associated to known gravity signals. For the drift estimation, all known gravity effects must be removed from the raw gravity data according to

$$\delta g_{\mathrm{SG\_res}} = \delta g_{\mathrm{raw}} - \mathrm{ET} - \delta g_{\mathrm{air}} - \delta g_{\mathrm{ol}} - \delta g_{\mathrm{hy}} - \delta g_{\mathrm{pol}} \tag{40}$$

The drift is simulated by a first-order polynomial $dr(t) = a_0 + a_1 \times t$ and the drift parameters a_0 and a_1 are determined by a linear fit of $\delta g_{\mathrm{SG\_res}}$. Because of some uncertainties in removing all gravity effects, the drift estimation can contain a real gravity signal too. To reduce the influence of insufficiently removed gravity effects, e.g. seasonal gravity effects from the hydrosphere, the drift estimate from a time series longer than 1 year gives a more realistic result.

More accurate, the drift estimation is realized by comparison with absolute gravity measurements for an interval of about 1 year.

Figure 10.14 g shows an example for the drift estimation at Metsahovi station. The calculated drift rate is 4.2 μGal/year.

10.7 Analysis of Surface Gravity Effects

For separating the different gravity effects, special analysis methods have been developed using gravity as well as additional meteorological and hydrological data. These analysis methods are based on different models for Earth and ocean tides, gravity variations induced by the atmosphere and hydrosphere, etc. Before analysing one specific effect such as Earth tides, seismic normal modes, core modes, nearly diurnal-free wobble (NDFW), or Chandler wobble, the disturbing signals must be removed.

One disturbing part consists of accelerations (vibrations) usually considered as noise (seismic, industrial and ocean noise), which can be reduced by low- or band-pass filtering (if the particular frequencies are known). Supplementary instrumental effects (drift, offsets and instrumental noise) superimpose the gravimeter signal. Most of these effects can be removed from the data. Another part usually treated as disturbing signals are environmental influences. To remove them, they have to be modelled and hence the environmental parameters of the atmosphere and

hydrosphere (e.g. atmospheric pressure, groundwater level, soil moisture) must be measured precisely.

Before analysis of a specific gravity effect (e.g. Earth tides), the gravity effects caused by the atmosphere, hydrosphere and the ocean are normally removed from the data as much as possible.

10.7.1 Pre-processing

The SG gravity data are recorded as raw data (unit: Volts) with a sampling rate of usually 1 s. Before analysis the data must be calibrated and exempt from spikes, offsets and drift. Short data gaps up to a few days can be filled by theoretical tides. The reasons for offsets are strong earthquakes, refilling of liquid helium (older instruments), instrumental causes, etc. Spikes result from high-frequency disturbances caused by different natural or man-made events and instrumental effects.

The pre-processing of the gravity data can be divided into different steps:

- Calibration of the data
- Phase (time) shift correction
- Computation of the preliminary gravity residuals
- Removing of spikes, steps and interfering signals
- Filling of short data gaps by theoretical tides
- Decimation to the required sampling rate for data analysis (e.g. 1 min for the free oscillation of the Earth or 1 h for tidal analysis)
- Drift correction

The calibration is done by multiplication of the data with the prior determined calibration factor (e.g. –68.945 μGal/V). Furthermore, according to the time shift determination (e.g. 8.7 s) the correct time must be assigned to the data.

Depending on the data quality, spikes larger than about 0.2 μGal (depending on the noise level of the data) and steps in the microgalileos range that do not have their origin in atmosphere or groundwater level-induced gravity variations are removed. Of special importance is the correction of steps in the raw data. This must be carried out with great care because steps in the data series influence the long-term result directly. Short data gaps can be filled by theoretical Earth tides in order to have a continuous data series. The gravity residuals are then zero at the time of the gap.

Data filtering should be performed according to the period range of the gravity effect, which shall be investigated. The filters must be from type zero phase shift, to avoid a phase shift of the data. The damping of the filter must be adapted to the dynamic range of the data, which depends on the dynamic range of the analogue to digital converter (ADC) in the data acquisition system. For an ADC with a 24-bit resolution, the damping of the filter must be 140 dB according to

$$D = 20 \log(2^{b-1}) + 2, \tag{41}$$

where D = damping of the filter and b = bit resolution of the ADC.

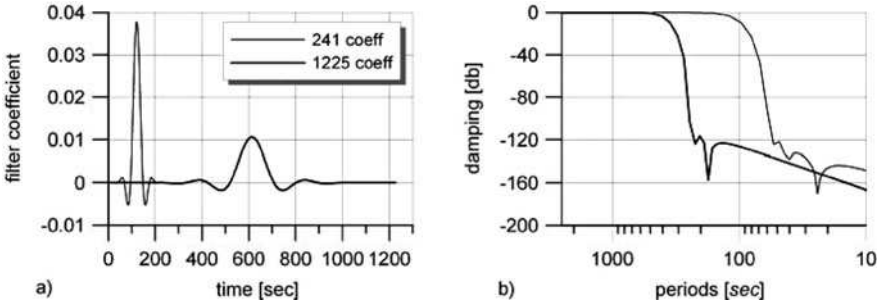


Fig. 10.31 FIR filter: (**a**) coefficients and (**b**) transfer function

A decimation filter should adequately cut off all frequencies higher than the Nyquist frequency, e.g. for 1 min data the Nyquist frequency $f_N = 1/(2\Delta t)$ yields $f_N = 0.00833$ Hz. The number of filter coefficients determines the length of the filter. Figure 10.31 shows an example of two typical filters with 241 and 1,225 coefficients and the respective transfer functions at a sampling rate of 1 s. The lengths of the filters are at 1 s sampling rate, 4 and 20.4 min, respectively. These filters can be used as decimation filters, e.g. from 1 s to 1 min. The filters can also be used at other sampling rates, and then the period in Fig. 10.31 is changing by multiplication with the sampling period. More decimation filters can be found on the GGP webpage (http://www.eas.slu.edu/GGP/ggpfilters.html).

The programs TSOFT (Van Camp and Vauterin 2005) and PRETERNA (Wenzel 1996) are tools capable of performing the above steps of the data pre-processing including data repair and filtering.

10.7.2 Earth Tides

The general goals of the tidal analysis consist of getting information (tidal parameters) of the response of the Earth due to the tidal acceleration. By using a multiple input–single output system this can be derived by a multiple regression model:

$$y(t) = \sum_{i=1}^{n} H_i\, A_i \cos(f_i t + \Phi_i + \Delta\Phi_i) + \sum_{m} R_m\, z_m(t) \tag{42}$$

with $y(t)$ = observations with the SG, H_i = amplitude factor, $\Delta\Phi_i$ = phase shift, A_i = amplitude and Φ_i = phase of the tidal wave with the frequency f_i, $z_m(t)$ = additional signal (e.g. atmospheric pressure), R_m regression parameter for the additional signal.

The unknown parameters H_i, $\Delta\Phi_i$ and R_m are normally determined by least square fit of the observations (Wenzel 1995c).

The recorded and pre-processed data series of the SG $\delta g_{\text{pre}}(t)$ consists not only of the tidal gravity signal ET(t) due to the tidal acceleration in a fixed point P. It also contains further information:

$$\delta g_{\text{pre}}(t) = \text{ET}(t) + \text{dr}(t) + \text{adm}\, a(t) + \text{res}(t), \tag{43}$$

where term dr(t) describes the drift of the gravimeter; term $a(t)$ is a time series with meteorological or hydrological data and adm describes the influence of this additional channel onto the gravity measurement. Further signals and measurement errors are combined in term res(t), the gravity residuals.

An analysis program optionally corrects the observed signal $\delta g_{\text{pre}}(t)$ by eliminating the drift series and the influence of the meteorological and hydrological signals. A comparison between the theoretical tidal acceleration (cf. Sect. 10.6.1) and the observed tidal gravity signal is used to estimate a set of tidal parameters (amplitude factor δ and phase lead κ) for the station. The tidal parameters amplitude factor and phase lead cannot be determined for each wave noted in the tidal potential catalogue. Following the Rayleigh criterion the waves of the used tidal potential catalogue are stacked together to wave groups. For each of these groups the tidal parameters, amplitude factor and phase lead are estimated. The Earth tides are the largest signal in the gravity recordings.

For the Earth tide analysis, three programs ANALYZE (Wenzel 1996), VAV (Venedikov et al. 2001) and BAYTYP-G (Tamura 1990) are in use. All three programs deliver results within very similar error bars (Dierks and Neumeyer 2002). The ANALYZE program is based on the least squares method using different tidal potential catalogues. The most used very accurate catalogue is the HW95 with a resolution of 1 nGal (Hartmann and Wenzel 1995a, b).

10.7.2.1 Program ANALYZE

The program is based on a method developed by Chojnicki (1973) and improved by Schüller (1976) and Wenzel (1996). A least squares adjustment is used to estimate the tidal parameters, the meteorological and hydrological regression parameters, the pole tide regression parameters and the Tschebyscheff-polynomial bias parameters for drift determination. The amount of data is nearly unlimited. Every kind of earth tide data (gravity, strain, tilt and displacement) and up to eight channels with meteorological and hydrological data can be analysed. The user can determine the range of up to 85 tidal wave groups. One tidal potential catalogue out of seven including the very accurate from Hartmann and Wenzel (1995a, b) can be chosen to calculate the theoretical tidal signal. The model used for least square adjustment is

$$\ell\,(t) + v\,(t) = \sum_{j=1}^{q} (\hat{X}_j\, \text{CO}_j + \hat{Y}_j\, \text{SI}_j) + \sum_k \hat{D}_k\, T_k(t_n) + \sum_m \hat{R}_m\, z_{m(t)} \tag{44}$$

with $\ell(t)$ = observed gravity signal, $v(t)$ = corrections of the observations, X_i , Y_i = linear form of the unknown tidal parameters H_i (amplitude factor) and $\Delta\Omega_i$

(phase lead) for each wave group j with

$$X_i = H_i \cos(\Delta\Phi_i), \tag{45}$$

$$Y_i = H_i \sin(\Delta\Phi_i), \tag{46}$$

where CO_j, SI_j = factors of theoretical tidal parameters A_j (amplitude) and Φ_j (phase) for each wave i in the wave group j, starting with wave a_i and ending with wave e_i:

$$\begin{aligned} CO_j &= \sum_{i=a_i}^{e_i} H_i^* A_i \cos(2\pi f_i t + \Phi_i), \\ SI_j &= \sum_{i=a_i}^{e_i} H_i^* A_i \sin(2\pi f_i t + \Phi_i) \end{aligned} \tag{47}$$

with H_i^* = amplification factor from digital high-pass filter (equal to 1 if the drift is approximated by polynomials), D_k, T_k = coefficients (D_k) of Tschebyscheff polynomials T_k and degree k and R_m, z_m = regression coefficients (R_m) of additional channel number m (z_m).

A possible drift in the data is eliminated by high-pass filtering or is approximated by Tschebyscheff polynomials (T_k) whose coefficients (D_k) are also estimated in the least square adjustment. The filter coefficients for different numerical digital filters are included in the ETERNA package (Wenzel 1996). However the method of high-pass filtering can only be used when no long periodic waves shall be determined. Together with the analysis of long periodic waves the drift can be eliminated by an approximation through Tschebyscheff polynomials.

The influence of the air pressure data (or other meteorological or hydrological signals $z_m(t)$) onto the gravity measurement is determined by a linear regression. In the case of high-pass filtering the air pressure data are filtered too and the regression is computed with the filtered data.

The accuracy of each parameter is estimated in the least square adjustment in the form of standard deviations. The standard deviations of the tidal parameters are too optimistic and therefore corrected. They are multiplied by a factor which is derived from the spectrum of the gravity residuals for the different frequency bands.

10.7.2.2 Program BAYTYP-G

This program is based on a method called Bayesian prediction, developed by Harrison and Stevens (1976). The method has been modified for the use with Earth tide data. All kinds of Earth tide data can be analysed, three additional channels with meteorological or hydrological data are possible. The arrangement of the tidal wave groups is done automatically depending on the length of the time series, but the user can change the wave group boundaries by editing the corresponding file (Tamura 1990; Tamura et al. 1991). The tidal potential catalogues from Tamura (1987) and others can be used.

Tidal parameters drift and meteorological parameters are estimated through an iterative method similar to least square adjustment by minimizing the function in (Tamura 1990):

$$\begin{aligned}\sum_{i=1}^{n}\left[y_i-\sum_{m=1}^{M}(A_mC_{mj}+B_mS_{mj})-d_i-\sum_{k=0}^{k_{\max}}b_k\,x_{i-k\Delta t}\right]^2\\ +D^2\sum_{i=1}^{n}\left[d_i-2d_{i-1}+d_{i-2}\right]^2\\ +\mathrm{WEIGHT}^2\sum_{m=2}^{M}\left((A_m-A_{m-1})^2+(B_m-B_{m-1})^2\right),\end{aligned} \tag{48}$$

where A_m and B_m are the linear expressions of the unknown amplitude factor and phase lead for each m of the M wave groups, respectively. C_{mj} and S_{mj} are computed from the tidal potential catalogue using all j waves contained in the mth wave group. This tidal part is subtracted from each observation y_i ($\delta g_{\mathrm{pre}}(t)$) together with the drift value d_i and the term describing the influence of additional channels $x(t)$ onto the measurement (cf. (7.2.2c)). D and WEIGHT are called hyper-parameters and can be defined in the parameter file.

The second line of (48) is used for drift computation. Within this program the drift is not approximated by low-degree polynomials. Here the drift is computed separately in each data point. The drift behaviour is characterized by the formula

$$d_i = 2d_{i-1} - 2d_{i-2} + u_i, \tag{49}$$

where u_i is the stochastic part denoting a white noise sequence, d_i is the drift value at the current data point and d_{i-1} and d_{i-2} are the drift values in the two previous data points. The hyper-parameter D can be used to fit the drift model to the data. A large value for parameter D selects an almost linear drift model; a small value leads to a drift model bending close to the data.

A similar possibility is given with hyper-parameter WEIGHT in the third line of (48). Here the variability of the tidal parameters can be chosen. However this option is only useful if too many tidal parameters shall be estimated from too poor data. The influence onto gravity measurement is computed by regression for maximum three additional signals.

Within the iterative search the hyper-parameter D is adjusted to get the best combination between parameters, measured data and tidal parameters. At the end of each turn an ABIC value (ABIC = AKAIKE Bayesian Information Criterion) is computed. The solution with the smallest ABIC value is the final one where data, parameter and drift fit each other best. This ABIC value is also the accuracy statement. A standard deviation is computed, but following the author of this program, it is simply derived from the ABIC value. So this standard deviation is not directly comparable to standard deviations from the other programs.

10.7.2.3 Program VAV

The program VAV is based on a method called MV66 (Venedikov 1966a, b) and is an improvement of the program NSV (Venedikov et al. 1997). The data file can be adopted from program ANALYZE, but program VAV has its own format for data files and uses own input files for tidal wave grouping and parameter settings. The wave group arrangement is done automatically depending on the length of the data set. Also a grouping variant can be chosen from a special input file.

The fundamental idea of the program NSV (Venedikov et al. 1997) is a filtering of the original data containing an elimination of the drift and the separation into several pairs of series (step 1). Each pair contains signals from one main tidal constituent in diurnal, semidiurnal and terdiurnal band. The unknown tidal parameters for each tidal constituent are determined simultaneously (step 2). This leads also to a frequency-dependent accuracy estimate (AKAIKE Information Criterion (AIC value) and standard deviation). Both steps are also contained in program VAV but the separation in step 1 is not restricted to main tidal constituents. Here a wide spectrum of frequencies can be chosen by the user. Step 2 is using all the separated tidal constituents in a single least square adjustment. An improvement of program VAV is the possibility to use data with different sampling rates and with several gaps in the same run without the need for interpolation (Venedikov et al. 2001).

The observed data sets are divided into N intervals y_i of equal length. Each set contains n data points. n differs between the intervals, if the data are unequally spaced. Tidal parameters and air pressure regression coefficient are determined in a least square fit together with the drift polynomial coefficients.

Besides the standard deviation from the least square adjustment, an AIC value is computed. This value is used to compare different solutions for the same data set with different parameter settings. The solution with the smallest AIC value is the best one.

10.7.2.4 Analysis Results

The high-precision estimation of the gravitational amplitude factor δ and phase κ for the different partial tidal waves, which represent the response of the Earth body to tidal forces are mainly used for

- Earth tide reduction for relative and absolute gravity and other precise measurements like satellite positioning, GPS-, laser-, radio-interferometric methods
- Investigation of Earth tide models
- Investigation of changes of the tidal parameters
- Determination of the resonance effect of the liquid core (nearly diurnal-free wobble)
- Verification of the Love numbers h and k by gravity (SG) and GPS, VLBI or other measurements
- Verification of global and regional ocean tide models (Baker and Boss 2003)

The accuracy of the tidal parameters depends on the SG calibration accuracy and the reduction of the other gravity effects in this period range (atmosphere, hydrology and ocean). With a calibration accuracy of 0.2% the tidal factors can be determined to $\Delta\delta = \pm 0.002$. The time shift of the SG can be determined with an accuracy better than 1 s. From this follows that the phase shift can be determined better than 0.01° (calculated for diurnal tidal waves).

Figure 10.32 shows the determined tidal parameters calculated from SG data (March 2000 to December 2004) of the SU station. For comparison the WD model (white columns) and the observed amplitude factors (grey columns) are presented. The deviations from the model can be seen clearly. One reason for the deviations is the influence of the ocean loading. Therefore the ocean loading correction has been calculated with LOAD97 and the co-tidal map from model FES2002 for the diurnal partial tides Q1, O1, P1 and K1 and the semidiurnal partial tides N2, M2, S2 and K2. The black columns show the ocean loading-corrected amplitude factors. One can see that the ocean loading-corrected amplitude factors come closer to the model values for the semidiurnal waves N2, M2, S2, K2 and the diurnal waves P1 and K1. For the diurnal waves Q1 and O1 the ocean loading-corrected amplitude factors depart from the model values. The reason for this behaviour is the ocean loading model also shown in Ducarme et al. (2002).

The model phase is zero. Larger deviations from the model phase show the semidiurnal waves 2N2, N2, M2, L2, S2, K2 and the diurnal wave S1. The ocean-corrected phase leads for the diurnal waves N2, M2, S2 and K2 give a good improvement close to zero (observed values near 5° phase lead). The ocean-corrected phase lead for the diurnal waves Q1, O1, P1 and K1 becomes larger than the uncorrected value. The strong deviation (δ and κ) of the S1 wave from the model is mainly caused by the influence of the daily variations of the atmospheric pressure,

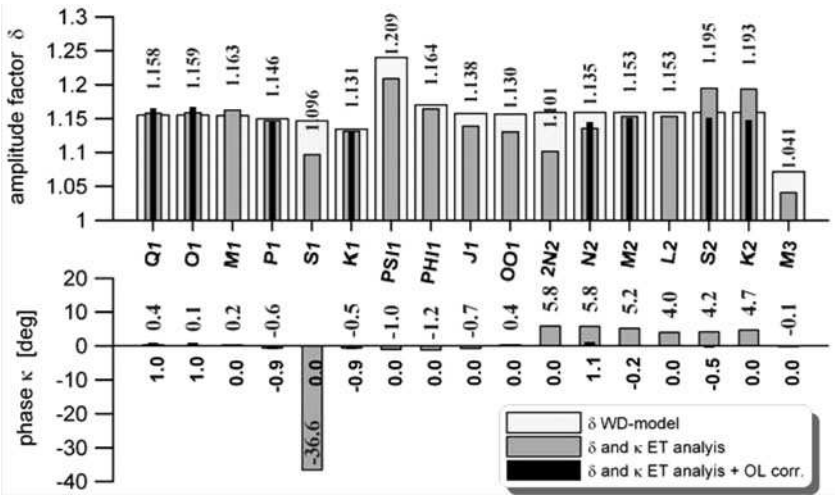


Fig. 10.32 Earth tide parameters δ and κ for the SU station, South Africa. *White columns*: WD model parameter δ; *grey columns*: calculated parameters δ and κ; *black columns*: ocean loading (model FES2002) corrected parameters δ and κ

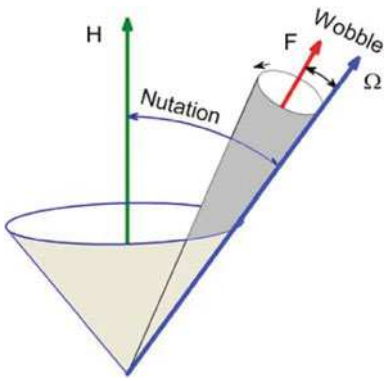


Fig. 10.33 Nearly diurnal-free wobble (NDFW)

which are not corrected completely. More analysis results from the GGP network can be found in Ducarme et al. (2002, 2004) and Sun et al. (2003).

10.7.3 Nearly Diurnal-Free Wobble

The nearly diurnal-free wobble (NDFW) is a small retrograde motion of the Earth's instantaneous rotation axis Ω about the figure axis *F* with a period close to 1 sidereal day, connected with the free core nutation (FCN), a retrograde motion of the rotation axis in space with respect to the angular momentum axis *H* with a period of around 430 sidereal days (Fig. 10.33). The figure axis of the mantle and the instantaneous rotation axis of the inner core are misaligned due to the tidal forcing. The mantle tries to impose its shape on the core. Restoring pressure torques arise at the core mantle boundary which try to realign the two axes. As a result, the Earth reacts as a gyroscope with a nutation in space and a wobble with respect to the Earth figure axis (Neuberg 1987).

The NDFW resonantly amplifies nearly diurnal tides and annual and semi-annual nutation. The tides which are mostly affected are P1, PI1, K1 PSI1 S1 and PHI1. Their amplitudes are influenced, because their frequencies are close to the eigenfrequency of the NDFW. Figure 10.34 shows the amplitude factors of the WD model around the eigenfrequency.

The eigenfrequency of the NDFW can be determined according to analytic solutions (Florsch et al. 1994; Florsch and Hinderer 2000; Lei and Xu 2002) or by least square fit estimation according to the model of a damped harmonic oscillator (Neuberg et al. 1987; Defraigne and Dehant 1994; Sun et al. 2002a; Ducarme et al. 2002). According to Defraigne and Dehant (1994) the complex amplitude factor can be expressed by

$$\delta(\sigma) = \delta_0 + \frac{\overline{A}}{\sigma - \overline{\sigma}_{\mathrm{nd}}} \tag{50}$$

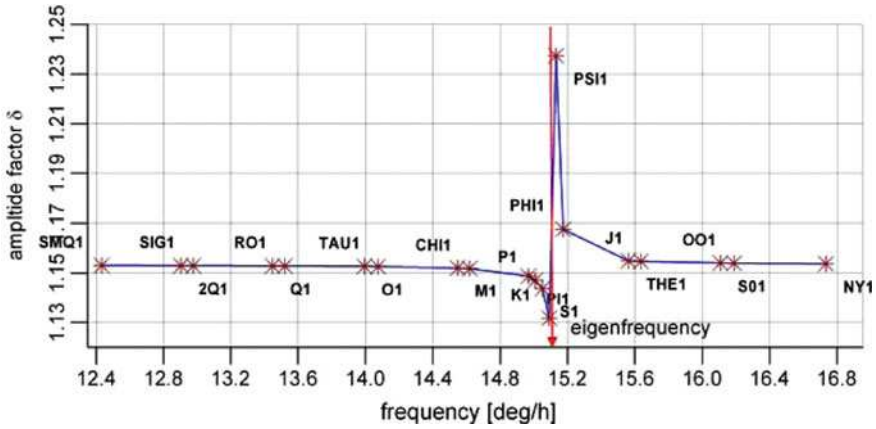


Fig. 10.34 Diurnal tidal waves influenced by the NDFW (amplitude factors from WD model)

with $\delta_0 = 1 + h_0 - \frac{3}{2}k_0$, where h_0 and k_0 are the static Love numbers, $\overline{\sigma}_{\mathrm{nd}}$ is the eigenfrequency of the NDFW, σ the frequency of the tidal wave close to $\overline{\sigma}_{\mathrm{nd}}$ with the amplitude factor $\delta(\sigma)$ and $\overline{A}$ the resonance strength depending on the geometric shape of the Earth and the rheological properties of the Earth mantle. With the assumption that the tidal wave O1 is not influenced by the resonance, because the O1 frequency is too far away from the resonance frequency, the tidal gravimetric factor δO1 is then subtracted from (50). This results in the model for determination of the complex number $\overline{\sigma}_{\mathrm{nd}} = \mathrm{Re}(\sigma_{\mathrm{nd}}) + i \times \mathrm{I}m(\sigma_{\mathrm{nd}})$ by least square adjustment of the parameters $\overline{A}$ and $\overline{\lambda}$nd:

$$\delta(\sigma) - \delta(\sigma_{O1}) = \left(\frac{\overline{A}}{\sigma - \overline{\sigma}_{\mathrm{nd}}}\right) - \left(\frac{\overline{A}(O1)}{\sigma(O1) - \overline{\sigma}_{\mathrm{nd}}}\right). \tag{51}$$

To solve (51) the Marquardt algorithm of linearized iteration can be used to minimize the error function χ^2:

$$\chi^2 = \sum_{\lambda,j} \omega(\sigma,j)\left[\delta(\sigma,j) - \delta(\lambda_{O1},j) - \left(\frac{\overline{A}_j}{\sigma - \overline{\sigma}_{\mathrm{nd}}} - \frac{\overline{A}(O1)}{\sigma(O1) - \overline{\sigma}_{\mathrm{nd}}}\right)\right]^2 \tag{52}$$

with $\omega(\sigma, j) = \frac{1}{\varepsilon(\sigma,j)}$ the weight function, where $\varepsilon(\sigma,j)$ is the standard deviation of the amplitude factor of the tidal wave with the frequency σ at the jth station (Defraigne and Dehant 1994; Xu et al. 2002).

The eigenperiod of the FCN T_{FCN} and the quality factor Q can be determined by (53) and (54):

$$T_{\mathrm{FCN}} = \frac{1}{\mathrm{Re}(\sigma_{\mathrm{nd}}) + \Omega}, \tag{53}$$

where $\Omega \sim 1$ cpsd (cycle per sidereal day) is the sidereal frequency of the Earth's rotation and

$$Q = \frac{\mathrm{Re}(\sigma_{\mathrm{nd}})}{2\ \mathrm{Im}(\sigma_{\mathrm{nd}})}. \tag{54}$$

For selection of the tidal waves for determination of the eigenfrequency of the NDFW the following criteria have to be considered. The tidal waves should be close to the eigenfrequency (highly affected by the resonance) and the tidal parameters must be determined with high accuracy. This means that the tidal waves must be corrected for the ocean loading effect (cf. Sect. 10.6.4. 1) and the tidal wave S1, which is strongly affected by the solar-heated atmospheric tide, should not be considered.

Good results could be achieved with the tidal waves O1 (very well determined tidal parameters), K1 (close to the resonance and well determined) and PSI1 (closest wave to the resonance, but because of the small amplitude not very well determined).

With the GGP network, using the stacking method (Neuberg et al. 1987), the FCN period has been determined, e.g. by Sato et al. (2004) to 429.7 ±1.4 sidereal days or Ducarme et al. (2007) to 429.7 ±2.4 sidereal days. It is close to the MAT01 model (Mathews et al. 2002) of 430 sidereal days.

10.7.4 Polar Motion

The gravity effect of the polar motion can be measured directly with the SG and compared with the polar motion derived from the IERS data. Some of the SG stations deliver very similar results like the IERS observations. The determined gravimetric amplitude factor δ_{Pol} ranges from 1.1 to 1.18. For some stations it is close to the WD model value of 1.16 (Hinderer and Crossley 2000; Harnisch and Harnisch 2006).

For measuring the polar motion all the other gravity effects must be removed from the SG data according to

$$\mathrm{PT}(t) = \delta g_{\mathrm{raw}}(t) - \mathrm{ET}(t) - \delta g_{\mathrm{air}}(t) - \delta g_{\mathrm{ol}}(t) - \delta g_{\mathrm{hy}}(t) - \mathrm{dr}(t). \tag{55}$$

Figure 10.35 shows an example of the polar motion measured with the SG at Sutherland site, South Africa, in comparison with the calculation from IERS data.

The polar motion is used for the determination of the orientation of the Earth's rotation axis, the separation of the annual and Chandler wobble part, verification of the polar motion measured with space techniques, determination of the gravimetric amplitude factor δ_{Pol}, etc.

10.7.5 Free Oscillation of the Earth

By a large earthquake the Earth is set into its natural oscillations (Earth's free oscillation). This oscillation is free and there are two independent types:

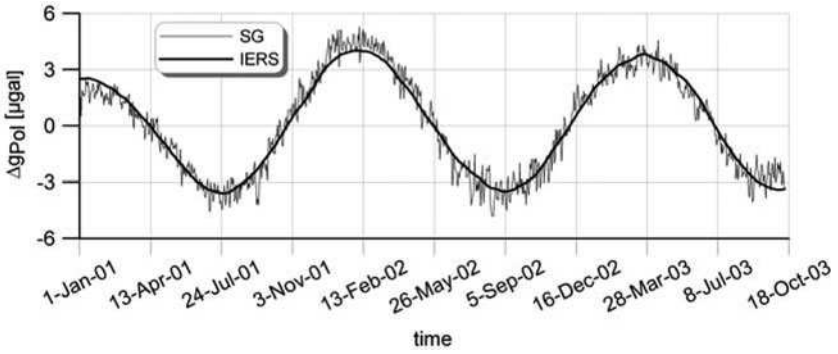


Fig. 10.35 Polar motion measured with the SG at SU site and calculated from IERS data

- Toroidal oscillations (T): the displacement for toroidal oscillations is always perpendicular to the radius vector. These oscillations involve only the crust and the mantle. They are equivalent to the Love waves.
- Spheroidal oscillations (S): the displacement for spheroidal oscillations has both radial and tangential components. They are equivalent to the Rayleigh waves.

Both types of oscillations have an infinite number of modes. The SG can clearly detect the S modes. Because of coupling effects on the rotating elliptical Earth, T modes can be detected by the SG too. Below 0.6 mHz the SG has the highest signal to noise ratio and it produces the best spectra and splitting of these modes (especially 0S2, 2S1 and 0S3). Based on this splitting the 3D density structure in the Earth's mantle and core can be investigated (Widmer-Schmidrig 2003). Figure 10.36 shows an example of the free oscillation after the Peru earthquake (latitude 16.14°S, longitude 73.312°W, depth 33 km) on 23 June 2001 at 20:33:14.14 with a magnitude of 8.4 recorded by the SG at SU station.

The data of this earthquake have been analysed for detection of the free oscillation modes of the Earth. For this purpose a data set of 96 h after the earthquake has been corrected for atmospheric pressure using the single admittance coefficient. An improvement of the atmospheric pressure correction can be achieved by special models adapted to this frequency range (Zürn 2002; Zürn and Wielandt 2007). After low-pass filtering (corner frequency 6 mHz) of the data, the mode spectrum has been calculated by using a Hanning window (Fig. 10.36). Above the spectrum the spheroidal modes are listed. Their frequencies are at the vertical grid lines. The spectrum shows the model modes up to the frequency of 0S10. In particular, the long-periodic modes 0S3, 2S1 and 0S2 are very well marked after this earthquake and less disturbed because of the low noise site. Further results can be found in Virtanen (1996), van Camp (1999), Nawa et al. (2000), Rosat et al. (2004, 2005) and Lei et al. (2005) where several earthquakes are analysed and the splitting of the modes is shown.

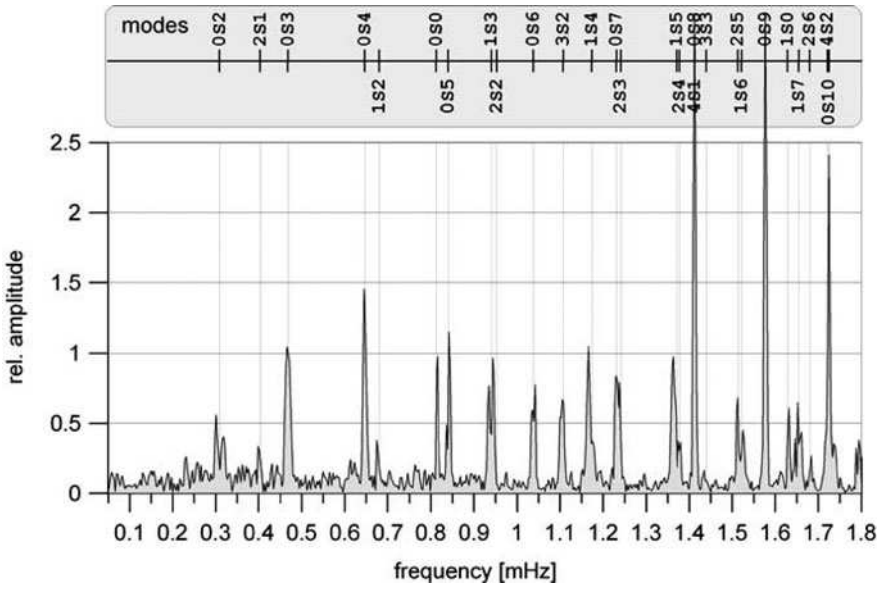


Fig. 10.36 Spheroidal free oscillation modes after the Peru earthquake on 23 June 2001

10.7.6 Translational Oscillations of the Inner Core (Slichter Triplet)

Slichter (1961) first pointed out that some of the Earth's free oscillations might consist principally of translational oscillations of the inner core about its equilibrium position at the mass centre of the Earth. Unlike Earth's other free oscillations, the restoring force on the inner core is primarily gravitational rather than elastic. Due to the Earth's rotation, a single mode is split into three different polarizations (the Slichter triplet): one along the axis of rotation and two in the equatorial plane (one prograde and one retrograde with respect to the sense of the Earth's rotation). Of all Earth's free oscillations, perhaps the translational motions of the solid inner core are the most exclusive. The signal of these oscillations at the Earth's surface will be a small periodic change in gravity below the nanogalileos level. Observational identification depends on the detection of all three distinct periods in time series from different sites.

The periods of the Slichter triplet are very sensitive to deep interior properties (e.g. the density jump at ICB (inner core boundary), viscosity of outer fluid core) and hence credible observational identification provides valuable information about the Earth's inner structure. Depending on the Earth model and the authors, different triplets for the modes were calculated and observed (Smylie et al. 1992, 2001; Rosat et al. 2004, 2006; Sun et al. 2004; Guo et al. 2006). The theoretical values of the frequencies of the Slichter modes are given, e.g. by Rogister (2003) for the Earth model 1066A (Gilbert and Dziewonski 1975) with 4.777, 5.229 and 5.813 cpd for

the prograde, axial and retrograde modes, and 4.006, 4.521 and 5.031 cpd for the PREM model.

Up to now the Slichter triplet has not been detected with high significance because their magnitude is close to the noise level of the SG data.

The SG network is the only configuration for detecting the Slichter triplet. The general method for detecting the modes consists in stacking of the data (Courtier et al. 2000) of different SG stations applying the product spectrum (Smylie et al. 1993) or frequency-sensitive algorithms.

10.7.7 Co-seismic Gravity Change

Gravity changes due to earthquakes have been observed with absolute gravimeters (e.g. Tanaka et al. 2001) and superconducting gravimeters (Goodkind 1999 and Imanishi et al. 2004). Tanaka et al. (2001) observed gravity changes of 6 μGal after an earthquake of magnitude 6.1 with an absolute gravimeter. Goodkind (1999) analysed SG gravity changes at Fairbanks, Alaska, corresponding to an earthquake within 500 km distance to the SG. He analysed a decreasing of gravity before the earthquake (16 May 1993, magnitude 4.0, depth 106 km, distance 126 km) and arising after the earthquake. These measured gravity changes accounted for some microgalileos, but were in contradiction to the straightforward models of earthquakes (Lambert and Bower 1991). The large gravity changes could be caused by gravity variations of other sources. Nevertheless, this study gave input to further studies. Imanishi et al. (2004) observed co-seismic gravity changes from the Tokachi-oki earthquake on 25 September 2003 at 19:50:08 with a magnitude of 8 at location $\phi = 41.78°$N, $\lambda = 144.079°$E close to the Hokkaido Island at a depth of 32 km. Co-seismic gravity changes have been recorded at the SG stations of Essashi (0.58 μGal, distance 3.4°), Matsushiro (0.1 μGal, distance 6.9°) and Kyoto (0.07 μGal, distance 9.4°). These results are in good agreement with the calculated results of dislocation models.

Generally, we have to consider that the sphere (prove mass) of the SG is partly out of feedback and recording range during strong earthquakes. The measurement range of the SG covers about ±900 μGal. The automatic resetting of the sphere into range may also cause an offset. This is the uncertainty in interpreting co-seismic gravity changes measured with SGs. On the other hand, the analysis period should not be too large, in order to reduce an interaction with gravity variations of other sources.

In Kim et al. (2009) four earthquakes within a distance of about 100 km to the MG station were analysed. During the largest earthquake EQ1 (20 January 2007 at 11:56:33 with a magnitude of 4.8 at location $\phi = 37.68°$N, $\lambda = 128.59°$E with a depth of 13.1 km) the maximal seismic signal was 811 μGal, close to the end of scale but still in feedback. The signals of the other earthquakes were within feedback too. Therefore, we can exclude an instrumental effect caused by out of feedback of the sphere as reason for the gravity change during the earthquake.

The recorded raw gravity data of the earthquakes (1 s sampling rate) were reduced for theoretical Earth tides, air pressure-induced gravity, ocean loading effect, polar motion and the linear drift. For detection of a co-seismic gravity change, an offset in the data close to the event time of the earthquake, a data set for about 1 h before and after the earthquake was selected. The data were divided into two blocks. Block 1 starts 1 h before the earthquake and ends 1 sample (1 s) before recording of the earthquake (event time t_0). Block 2 starts 2 min after the event time and ends after 1 h. The first 2 min of the recorded earthquake was not used for fitting because of the large amplitudes. After 2 min the amplitudes were ±2.5 μGal in maximum. The data of the two blocks were separately fitted by a quadratic function (56) to consider the different nonlinear trends before and after the earthquake:

$$y(t) = a + bt + c\,t^2. \tag{56}$$

Then the gravity of the fitted functions at event time t_0 for block 1, $y1(t_0)$, and block 2, $y2(t_0)$, was calculated. The co-seismic gravity change $\Delta g_{\text{co - seis}}$ was determined by

$$\Delta g_{\text{co-seis}} = y2(t_0) - y1(t_0). \tag{57}$$

The calculated gravity change is $\partial g_{\text{co-seis}} = 0.06$ μGal. Figure 10.37 shows the recorded earthquake EQ1 and the fitted curves before and after the earthquake. Because other effects can cause a gravity change, the atmospheric pressure and the groundwater level data were analysed for a step at and around the event time of the earthquake. A step could not be found in these data.

Of course, for confirmation of the observation results, modelling with a dislocation model should also be carried out. But this is an example where the sphere was

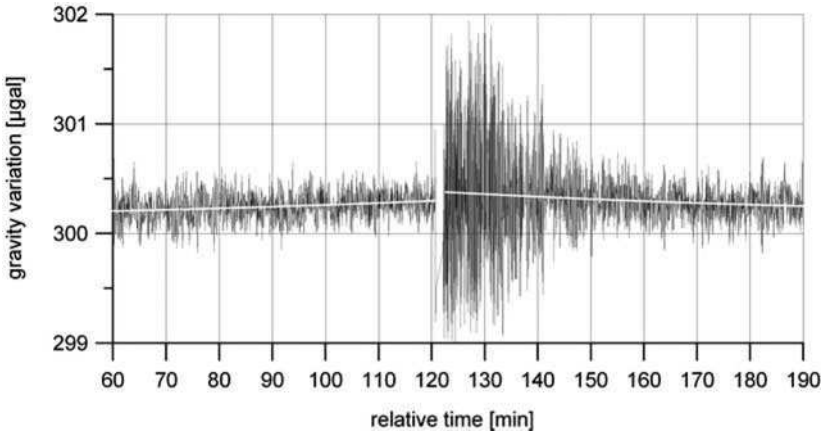


Fig. 10.37 Earthquake EQ1 recorded at MG station (*black*) and fitted quadratic function before and after the earthquake (*white*)

in feedback during the earthquake. A pre-seismic signal could not be detected for this earthquake.

10.7.8 Gravity Residuals

After reduction of the known gravity effects from the raw gravity data, the remaining part, the gravity residuals δg_{res} (58), cannot normally be associated to a specific gravity effect:

$$\delta g_{\mathrm{res}}(t) = \delta g_{\mathrm{raw}}(t) - \mathrm{ET}(t) - \mathrm{PT}(t) - \delta g_{\mathrm{air}}(t) - \delta g_{\mathrm{ol}}(t) - \delta g_{\mathrm{hy}}(t) - \mathrm{dr}(t). \quad (58)$$

If we do not subtract all the known gravity effects in (58), the resulting gravity residuals can be used for further investigation of these other gravity effects (e.g. pole tide, hydrology-induced gravity). In Fig. 10.38 an example is shown where all the known gravity effects are reduced except the effect of the global hydrology δg_{hy}.

From the gravity variations δg_{SG} monthly arithmetic means $\delta g m_{\mathrm{SG}}$ are calculated for comparison with GRACE or hydrological models-derived gravity variations (cf. Sect. 10.8.2).

For further analysis of the gravity residuals least square fit, cross-spectral analysis, product spectrum, power spectrum, frequency analysis (modified Fourier analysis), principal component analysis, etc. are in use.

Figure 10.39 shows an example (Kim et al. 2009) for detection of unknown frequencies in the gravity residuals δg_{res} at MunGyung station in Korea (September 2005 to January 2007) by means of the power spectrum. It was applied with a subdivision length of the gravity residual data of 18 days and an overlapping of 70%, and the data were multiplied with a Hanning window.

The power spectrum shows strong peaks at frequencies 1–7 and 11 cpd. The same peaks can be found in the power spectral density of the atmospheric pressure. These are the well-known peaks in gravity residuals (cf. Kroner 1997). They correspond to the harmonics of the solar-heated atmospheric wave S1. These waves could not

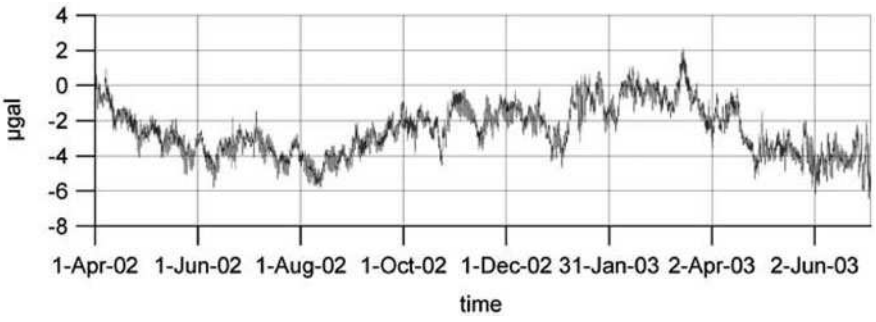


Fig. 10.38 Gravity residuals (Metsahovi station, Finland)

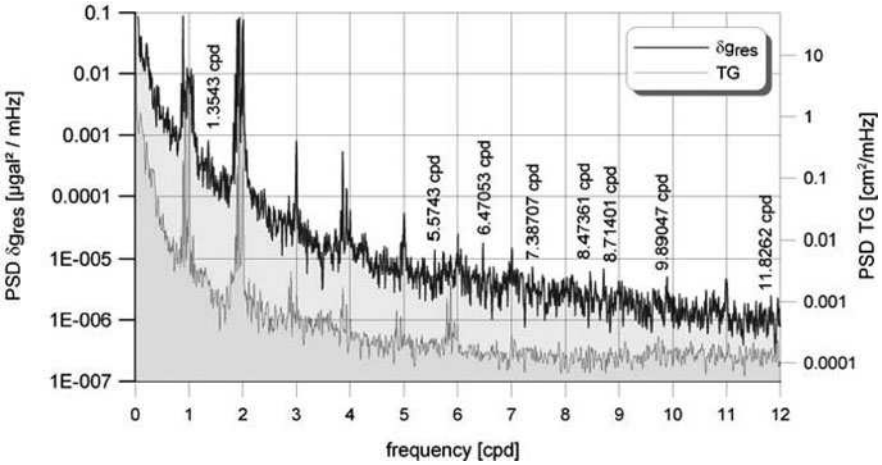


Fig. 10.39 Power spectrum of gravity residuals (September 2005 to January 2007) at MunGyung station, Korea

be taken into account by using the single admittance coefficient for calculation of the air pressure-induced gravity. Therefore, they remain in the gravity residuals. By using the frequency-dependent admittance these waves also cannot be removed completely from the gravity residuals (Crossley et al. 1995; Neumeyer 1995).

For checking frequencies which do not have their origin in the ocean loading, the power spectrum of tide gauge measurements (TG) from the Korean station East Hoopo (36° 40′ 28″N; 129° 27′ 20″E) with a distance to the MG station of 109 km was added. Local tide gauge measurements correlate well with the ocean loading and can improve the ocean loading correction compared to global ocean tide models (Neumeyer et al. 2005). The power spectrum of TG shows the diurnal (1 cpd) and semidiurnal (2 cpd) ocean tides which are not completely removed in the gravity residuals δg_{res}. Moreover, around 3 and 4 cpd and at larger frequencies are peaks in both spectra mainly caused by atmospheric and ocean loading. More detailed studies about this subject were done, e.g. by Anderson (1999), Boy et al. (2004) and Khan and Hoyer (2004).

Furthermore, there are several peaks in δg_{res} outside these frequencies, e.g. $f1 = 1.3543$ cpd, $f2 = 5.5743$ cpd, $f3 = 6.47053$ cpd, $f4 = 7.38707$ cpd, $f5 = 8.47461$ cpd, $f6 = 8.71401$ cpd, $f7 = 9.89047$ cpd and $f8 = 11.8262$ cpd. These peaks cannot be found in the spectra of the atmospheric pressure, groundwater level and ocean loading-induced gravity. The sources for these frequencies are unknown. They must be further investigated.

10.8 Combination of Ground (SG) and Space Techniques

The space techniques, e.g. GPS (Global Positioning System) and GRACE (Gravity Recovery and Climate Experiment) (Tapley and Reigber 2001), have achieved

a resolution at the millimetre level for the vertical surface displacement and in microgalileos level for the vertical gravity changes. This remarkable resolutions used for many applications in geodynamics can also be used for comparison with ground-based (SG) measurements with the goal of validating different techniques.

10.8.1 Combination of SG and GPS Measurements

The GPS technique can achieve a positioning precision in the millimetre range for displacement variation (altitude variation) of a point on the Earth's surface, e.g. for the SG site. To get this precision, one needs a GPS network of at least two stations. One station is the reference station, which is assumed as fixed and error free, and the coordinate variations of the other station, which is located at the SG site can be determined. Adding more reference stations can improve the precision of the result. The coordinates of the reference station must be known before processing to better than 1 cm. This can be achieved by using data of the IGS (International GPS Service) GPS network or by using IGS stations as reference stations. The distance between the GPS stations should be between 100 and 800 km approximately. The minimal observation time can be a few days only but a parallel registration to the SG allows the detection of long-term variations.

One can get the precision in the millimetre range with a static GPS solution only. If a dynamic solution is applied, the precision will be only in the centimetre range (Xu 2003). The comparison between SG and GPS measurements offers the separation of gravity variations caused by vertical displacements and the associated mass redistributions. In Zerbini et al. (2001) good correlations between gravity measured with the SG and GPS measured altitude variations can be found.

10.8.2 Comparison of SG, GRACE and Hydrological Models-Derived Gravity Variations

One objective of the satellite gravity mission GRACE is the recovery of temporal Earth gravity field variations. The results of GRACE data evaluation show a gravity resolution in the microgalileos range at a half wavelength $\lambda/2$ spatial resolution of about 1,500 km for a temporal resolution of 1 month (Schmidt et al. 2005). Because of this remarkable gravity resolution, the comparison and validation of satellite-derived temporal gravity variations with ground gravity measurements is of fundamental interest. For a satellite the gravity resolution is connected with the spatial resolution. With higher degree ℓ, the gravity resolution decreases and the spatial resolution becomes finer. For comparison with SG measurements monthly solutions up to $\ell_{max} = 20$ (gravity resolution ~ 2.5 μGal and $\lambda/2$ ~ 1,000 km) can be applied. Since the time variations contained in the GRACE solutions range from 1 month to the life time of GRACE, terrestrial gravity measurements must have a long-term stability, which can be fulfilled by SGs only.

A combination of satellite-derived with ground-measured temporal gravity variations requires representative data sets with the same sources of gravitation and comparable spatial resolutions. This can be achieved by

- Reduction of the same gravity effects in both data sets (SG and GRACE) using the same models. Both data sets must represent the same sources of gravitation and the same spatial resolution. Therefore, all local gravity effects must be removed from the ground measurements.
- In contrast to the SG, the satellite is not coupled to the Earth's surface and therefore not sensitive to a vertical surface shift (altitude variation). This shift is described by the body Love numbers h_ℓ and the load Love numbers h'_ℓ. This fact must be considered for reducing the different gravity effects.
- Adaptation of the SG gravity variations to the spatial resolution of the satellite.
- Consideration of effects that are unique to each method.

Located on the Earth's surface, an SG measures, besides the gravitational mass attraction, the gravity effect due to elastic deformation (vertical surface shift) (e.g. Pick et al. 1973; Vanicek and Krakiwsky 1982) and the deformation potential (mass redistribution due to vertical surface shift). A satellite is not coupled to the Earth's surface and hence is only sensitive to the change in potential. It measures the mass redistributions only. The reductions of Earth and pole tide, as well as the loading effects of atmosphere and hydrosphere, are different for SG and GRACE. The body Love numbers h_ℓ and load Love numbers h'_ℓ, which describe the altitude variations, are only relevant for the SG data processing.

After reduction of known gravity effects, the spatial resolution for the remaining gravity variation is still different for SG (point measurements) and GRACE (used spatial resolution between 1,000 and 2,000 km). The SG measurements include all gravity variations from short to long-periodic spatial distribution. For comparison, only the gravity variations related to the spatial resolution of GRACE should be taken into account. The present way for adapting the remaining SG gravity variations to the spatial resolution of GRACE consists in removing the local gravity effects, mainly induced by local hydrology, from SG data.

10.8.2.1 Preparing of the Data Sets

Before comparing SG and GRACE satellite-derived gravity variations, the SG measurements must be reduced for the same gravity effects that have been applied in the GRACE data processing. These effects are

- Earth tides
- Pole tide
- Gravity variations induced by the atmosphere
- Ocean tidal loading

The same models are used for both sets of gravity variations to reduce these effects. The hydrosphere-induced gravity variations are not corrected in the monthly GRACE solutions, which are used for comparison. Therefore, they are also not reduced in the SG data.

In addition to the direct gravity field variations, the SG also measures the gravity changes due to the load-induced variations of the radial position of the SG, whereas the satellite-derived models naturally do not contain this effect. To compare the satellite-derived gravity variations with that measured by the SG, it is necessary to add to the gravity variations from GRACE (δg_{G}) the altitude-induced loading part (changing of the SG's vertical position) (δg_{load}) which is included in the SG data. It can be described by

$$\begin{aligned}&\delta g_{\mathrm{G+load}}(\varphi,\lambda) = \delta g_{\mathrm{G}} + \delta g_{\mathrm{load}} = \\ &\frac{GM}{R^2}\sum_{\ell=0}^{\ell_{\max}}(\ell+1-2h'_\ell)\sum_{m=0}^{\ell}\left[\delta\overline{C}^{\mathrm{G}}_{\ell m}\cos(m\lambda)+\delta\overline{S}^{\mathrm{G}}_{\ell m}\sin(m\lambda)\right]\overline{P}_{\ell m}(\sin\varphi),\end{aligned} \tag{59}$$

where φ, λ are the spherical geocentric coordinates of the computation point (longitude, latitude), R is the reference radius (mean equatorial radius of the Earth), GM is the gravitational constant times mass of the Earth, ℓ, m are degree and order of the spherical harmonics, respectively, $\ell_{\max}$ is chosen maximum degree in practical calculations (any natural number, $\ell_{\max} < \infty$) and h'_ℓ are degree-dependent load Love numbers (Farrell 1972; Zürn and Wilhelm 1984; Hinderer and Legros 1989). $\overline{P}_{\ell m}$ are the fully normalized Legendre functions and $\overline{C}_{\ell m}, \overline{S}_{\ell m}$ are the fully normalized Stokes' coefficients. Superscript G is related to spherical harmonic coefficients of GRACE gravity variations. More details can be found in Neumeyer et al. (2006).

Time series of the monthly GRACE-only models are developed, processed and archived in the Science Data System (SDS) (available at GFZ Information System and Data Center (ISDC) http://isdc.gfz-potsdam.de/grace or at the International Centre for Global Earth Models (ICGEM) http://icgem.gfz-potsdam.de/ICGEM/ICGEM.html.) shared between the Jet Propulsion Laboratory (JPL), the University of Texas, Center for Space Research (UTCSR) and the GeoForschungsZentrum Potsdam (GFZ). These time series of the monthly GRACE-only models are labelled as Release (RLnn). The Release number (nn) points out the improved processing and extended time series, e.g. RL03 consisting of 41 monthly models in the period from February 2003 to August 2006. Missing are the fields for June 2003 and January 2004 due to limitations in the amount of useable data in these 2 months. The available models are expanded up to degree and order 120. General aspects of the gravity recovery from GRACE can be found in Reigber et al. (2005), Schmidt et al. (2005) and Neumeyer et al. (2006).

As an example, the assigned GRACE values are taken from the monthly global gravity field solutions GFZ RL03, and the GRACE-derived gravity variations with added loading part $\delta g_{\mathrm{G+load}}$ (59) are calculated for the coordinates of the SG site Bad Homburg (Fig. 10.40). The spherical harmonic coefficients are low-pass filtered by

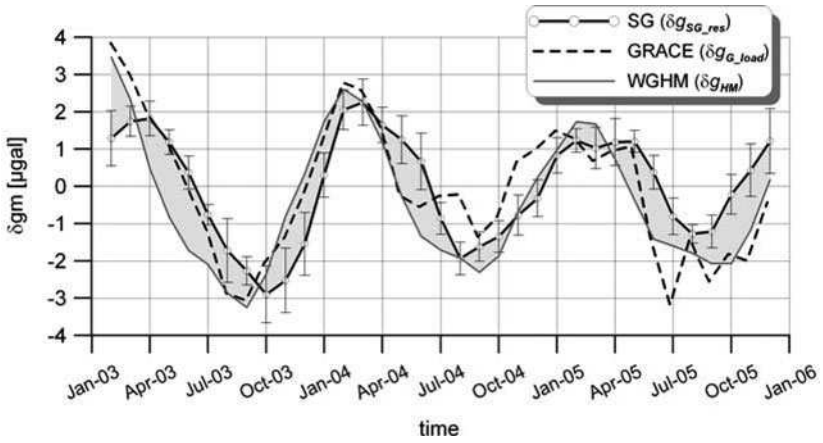


Fig. 10.40 Gravity variations derived from SG ($\delta g_{\mathrm{SG\_res}}$), GRACE ($\delta g_{\mathrm{G+load}}$) and hydrological model WGHM (δg_{HM}) for the site at Bad Homburg, Germany

Gaussian averaging in the space domain up to $\ell_{\mathrm{fl}} = 15$, which corresponds to a damping of the spherical harmonic coefficients by a Gaussian bell-shaped function (Jekeli 1981).

For comparison, the SG data were processed as follows. In a pre-processing procedure, spikes and steps due to instrumental and other perturbations such as earthquakes were carefully removed from the raw SG recordings. Spikes larger than 0.2 μGal and steps that do not have their origin in atmosphere or groundwater level-induced gravity variations were removed. Of special importance is the correction of steps in the raw data which are associated with instrumental problems (e.g. liquid helium transfer). This must be carried out with great care because steps in the data series influence the comparison result directly. Then, the data were low-pass filtered with a zero-phase shift filter (corner period 300 s) and reduced to 1 h sampling rate. From these pre-processed gravity data (δg_{raw}), which include gravity variations of different sources, the same gravity effects were subtracted as in the GRACE data processing.

The Earth tide reduction was performed with the WD model (Dehant 1987). The tidal parameters from this model were used for calculating the tidal gravity effect (ET) applied for semidiurnal to long-periodic constituents (tidal frequencies 3.190895–0.00248 cpd).

The gravity effect of the polar motion (PT) was calculated according to (39) based on IERS polar motion data.

For calculating the atmospheric pressure effect ($\delta g_{\mathrm{air}} = \delta g_{\mathrm{air\_attr}} + \delta g_{\mathrm{air\_def}}$), three-dimensional atmospheric pressure data from ECMWF were used. Because this gravity effect consists of both the attraction and the deformation parts, the calculation of the attraction term ($\delta g_{\mathrm{air\_attr}}$) was performed according to (27). The deformation term ($\delta g_{\mathrm{air\_def}}$) was calculated according to the Green's function for atmospheric loading (Merriam 1992; Sun 1995) due to (24).

Based on the global ocean tide model FES2002 (Lefevre et al. 2002; Le Provost et al. 2002), the ocean loading (δg_{ol}) for the main waves in semidiurnal, diurnal and long-periodic bands were calculated based on Francis and Mazzega (1990).

The local hydrology gravity effect (δg_{gwl}) was derived from groundwater level measurements in a borehole close to the SG site by the linear regression model (cf. Sect. 10.6.3.1). This method is a very simple one, because the reduction of the local hydrological gravity effect based on groundwater level variations near the SG station may also include a part of the global effect. The reduction of the gravity effect caused by the local hydrology possibly raises some problems. If the hydrology signal is a very local one, we must correct the effect. If the local hydrology effect contains components of the global hydrology, we would reduce the signal we want to compare. It is difficult to distinguish between local and global hydrological signals. Separating local from regional/global hydrological signals is still a challenge for interpreting temporal gravity variations.

The series of the GRACE monthly solutions are assumed as drift free. However, the SG has an almost linear instrumental drift, which can be calculated according to Sect. 10.6.6.

After reduction of all these gravity effects from the raw gravity data, the remaining part can be assumed to be mainly of mass changes in terrestrial water storage $\delta g_{\mathrm{SG\_res}}$:

$$\delta g_{\mathrm{SG\_res}}(t) = \delta g_{\mathrm{raw}}(t) - \mathrm{ET}(t) - \mathrm{PT}(t) - \delta g_{\mathrm{air}}(t) - \delta g_{\mathrm{ol}}(t) - \delta g_{\mathrm{gwl}}(t) - \mathrm{dr}(t) \quad (60)$$

From gravity variations $\delta g_{\mathrm{SG\_res}}$, monthly arithmetic means $\delta g m_{\mathrm{SG}}$ are calculated for comparison with GRACE and hydrological models-derived gravity variation.

The gravity variations δg_{HM} derived from global hydrology model WGHM (Döll et al. 2003) were calculated according to (33) up to $\ell_{fl} = 15$.

10.8.2.2 Comparing Results

One example of comparison is displayed in Fig. 10.40. It shows the gravity variations derived from SG ($\delta g m_{\mathrm{SG}}$), GRACE ($\delta g_{\mathrm{G\_load}}$) and hydrology model WGHM (δg_{HM}) for the Bad Homburg site. The error bars at the SG graph do not represent measurement errors; they show the variations of gravity within this month. All curves show a good agreement within their estimated error bars. The correlations between SG and GRACE, SG and WGHM, GRACE and WGHM are around 0.8.

In Neumeyer et al. (2008) it could be shown that the comparison between GRACE, SG and global hydrology model-derived gravity variations shows a noticeable agreement for many SG stations of the GGP network, except for SG underground stations. For these stations the modelling of the hydrological effect above the station is difficult and not very accurate.

The comparison results give suggestions for further investigations of each data series. They show furthermore that SG measurements of the GGP SG network can contribute to validate GRACE and further gravity satellite missions. In addition, field SG measurements should be carried out for further validation in areas with

large or very small gravity variations, e.g. the Amazon area, where seasonal gravity changes can be observed in the order of some 10 μGal or in the Atacama desert with a very small hydrological signal. The SG as well as the GRACE measurements can also contribute to improve the hydrological models.

More investigations on this subject can be found in Wahr et al. (2004), Crossley et al. (2005), Schmidt et al. (2006) and Hinderer et al. (2006)

10.9 Future Applications

Besides measuring the global gravity effects, the SG can also be used for measuring special regional and local gravity effects, because it is possible to remove all the other effects from the data, which can be well separated by special analysis and modelling methods. Some examples and proposals are

- Measuring of local hydrology signals for solving interesting hydrologic questions like area-averaged water storage, storage change, infiltration time, aquifer heterogeneity. The project of the University of Austin measures the hydrological signal with a transportable SG at different places at the Edwards Aquifer of Central Texas (Wilson et al. 2007).
- Volcano observations by combination of SG and GPS measurements for detecting mass transport (mass redistribution) and caused ground deformation through this. By combination of both measurements, a separation of mass transport (SG) and deformation (SG and GPS) is possible. The deformation can be determined with GPS stations in the millimetre range. A correlation of the combined signal (SG and GPS) with seismic, electromagnetic, chemical measurements, etc. can yield new information.
- Mass transport and crustal deformation in tectonic active zones.
- Investigation of the structure of the Earth's deep interior.
- Detection of silent earthquakes by a combination of SG, GPS and seismic measurements. For instance, in the Cascadian subduction zone, where displacements in the centimetre range on the Earth's surface have been observed (Rogers and Dragert 2003).

The present measuring and modelling precision including the separation of the different global and local gravity effects is sufficient for SG-based investigation of regional gravity effects such as secular gravity variations from postglacial rebounds, postseismic deformations, mass transport in tectonically active areas, hydrology signals, volcano activities, silent earthquakes.

Acknowledgement I am grateful to *Dr. R. Warburton,* GWR Instruments, San Diego, California, for providing detailed information about the SG and Figs. 10.2 and 10.3; *Dr. Franz Barthelmes* and *Dr. Sveto Petrovic,* GeoForschungsZentrum (GFZ), Potsdam, Germany, for the good cooperation during my active working time at GFZ, especially in the field of comparison of SG, GRACE and hydrological models; *Dipl. Ing. Olaf Dierks* for the good cooperation in SG data analysis during his active working time at GFZ Potsdam, Germany; *Dr. He Ping Sun*, Institute of

Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China; *Dr. Corinna Kroner* and *Dr Sveto Petrovic,* GFZ Potsdam, Germany; *Dr. Bruno Meurers*, Department of Meteorology and Geophysics, University of Vienna, Austria, for helpful comments and suggestions; *Prof Jeong Woo Kim*, Department of Geoinformation Engineering, Sejong University, Seoul, Korea and Department of Geomatics Engineering, University of Calgary, Calgary, Canada, for providing the gravity, atmospheric pressure and groundwater level data of the MunGyung SG station in Korea and good cooperation in analysis of these data; *Dr. Heikki Virtanen,* Finnish Geodetic Institute, for providing the gravity, atmospheric pressure and groundwater level data of the Metsahovi SG station in Finland and *Dr. Herbert Wilmes*, Bundesamt für Kartographie und Geodäsie, Frankfurt/M, Germany, for providing the gravity, atmospheric pressure and groundwater level data of the Bad Homburg SG station in Germany.

References

Anderson, O.B. (1999) Shallow water tides in the northwest European shelf region from TOPEX/POSEIDON altimetry. J. Geophys. Res., 104(C4), 7729–7774

Baker, T. and Boss, M. (2003) Validating Earth and ocean tides models using tidal gravity measurements. Geophys. J. Int., 152, 468–485

Banka, D. and Crossley, D. (1999) Noise levels of superconducting gravimeters at seismic frequencies. Geophys. J. Int., 139, 87–97

Bendat, J.S. and Piersol, A.G. (1986). Random Data-Analysis and Measurement Procedures. A Wiley-Interscience Publication, Wiley, New York, Chichester, Brisbane, Toronto, Singapore

Boy, J.P. and Chao, B.F. (2005) Precise evaluation of atmospheric loading effect on Earth time-variable gravity field. J. Geophys. Res., 110, B08412, doi: 10.1029/2002JB002333

Boy, J.P. and Hinderer, J. (2006) Study of the seasonal gravity signal in superconducting gravity data. J. Geodyn., 41, 227–233

Boy, J.P., Gegout, P. and Hinderer, J. (2002) Reduction of surface gravity data from global atmospheric pressure loading. Geophys. J. Int., 149, 534–545

Boy, J.P., Hinderer, J. and Ferhat, G. (2005) Gravity changes and crustal deformation due to hydrology loading. Geophys. Res. Abstr., 7, 07166

Boy, J.P., Hinderer, J. and Gegout, P. (1998) The effect of atmospheric loading on gravity. In: Ducarme, B. and Paquet, P. (eds) Proceedings of the 13th International Symposium on Earth Tides, Brussels, 1997, pp. 439–440

Boy, J.P., Llubles, M., Ray, R., Hinderer, J., Florsch, N., Rosat, S., Lyard, F. and Letelier, T. (2004) Non-linear oceanic tides observed by superconducting gravimeters in Europe. J. Geodyn., 38, 391–405

Chojnicki, T. (1973) Ein Verfahren zur Erdgezeitenanalyse in Anlehnung an das Prinzip der kleinsten Quadrate. Mitteilungen aus dem Institut für Theoretische Geodäsie der Universität Bonn, Nr. 15

Courtier, N., Ducarme, B., Goodkind, J., et al. (2000) Global superconducting gravimeter observations and the search for the translational modes of the inner core. Phys. Earth Planet. Int., 117, 3–20

Creutzfeldt, B., Günther, A., Klügel, T. and Wziontek, H. (2008) Simulating the influence of the water storage change on the superconducting gravimeter of the Geodetic Observatory Wettzell. Geophysics, 73(6), 95–104

Crossley, D. (2004) Preface of the global geodynamic project. J. Geodyn., 38(3–5), 225–236

Crossley, D., Hinderer, J. and Boy, J.P. (2005) Time variation of the European gravity field from superconducting gravimeters. Geophys. J. Int., 161, 257–264

Crossley, D., Hinderer, J., Casula, O., Francis, O., Hsu, H.T., Imanishi, Y., Jentzsch, G., Kääriäinen, J., Merriam, J., Meurers, B., Neumeyer, J., Richter, B., Sato, D., Shihuya, K. and van Dam, T. (1999) Network of superconducting gravimeters benefits a number of disciplines. EOS. Trans. Am. Geophys. Union, 80(11), 125–126

Crossley, D., Jensen, O.G. and Hinderer, J. (1995) Effective barometric admittance and gravity residuals. Phys. Earth Planet. Int., 90, 221–241

Crossley, D. and Su, X. (1998) Comprehensive analysis of 2 years of SG data from Table Mountain Colorado. Geophys. J. Int., 135, 835–844

Defraigne, P. and Dehant, V. (1994) Stacking gravity tide measurements and nutation observation in order to determine the complex eigenfrequency of the nearly diurnal free wobble. J. Geophys. Res., 99(B5), 9203–9213

Dehant, V. (1987) Tidal parameters for an inelastic Earth. Phys. Earth Planet. Int., 49, 97–116

Dehant, V., Defraigne, P. and Wahr, J. (1999) Tides for a convective Earth. J. Geophys. Res., 104(B1), 1035–1058

Dierks, O. and Neumeyer, J. (2002) Comparison of Earth tides analysis programs. Bull. Inf., Marees Terrestres, 135, 10669–10688

Döll, P., Kaspar, F. and Lehner, B. (2003) A global hydrological model for deriving water availability indicators: model tuning and validation. J. Hydrol., 270, 105–134

Ducarme, B., Sun, H.P. and Xu, J.Q. (2007) Determination of the free core nutation period from tidal gravity observations of the GGP superconducting gravimeter network. J. Geod., 81, 179–187

Ducarme, B., Sun, H.P. and Xu, Q. (2002) New investigations of tidal gravity results from the GGP network. Bull. Inf. Marees Terrestres, 136, 10761–10776

Ducarme, B., Venedikov, A.P., Arnoso, J. and Vieira, R. (2004) Determination of the long period tidal waves in the GGP superconducting gravity data. J. Geodyn., 38, 307–324

Etling, D. (2002) Theoretische Meteorologie: Eine Einführung, Springer, Heidelberg

Falk, R., Harnisch, M., Harnisch, G., Novak, I. and Richter, B. (2001) Calibration of the superconducting gravimeter SG 103, C023, CD029 and CD030. J. Geodetic Soc. Jpn., 47(1), 22–27

Fan, Y. and van den Dool, H. (2004) The CPC global monthly soil moisture data set at 1/2 degree resolution for 1948-present. J. Geophys. Res., 109, D10102, doi: 1029/2003JD004345

Farrell, W.E. (1972) Deformation of the Earth by surface loads. Rev. Geophys. Space Phys., 10, 761–797

Florsch, N., Chambat, F., Hinderer, J. and Legros, H. (1994) A simple method to retrieve the complex eigenfrequency of the Earth's nearly diurnal free wobble; application to the Strasbourg superconducting gravimeter. Geophys. J. Int., 116, 53–63

Florsch, N. and Hinderer, J. (2000) Bayesian estimation of the free core nutation parameters from the analysis of precise tidal gravity data. Phys. Earth Planet. Int., 117, 21–35

Francis, O. (1992) Interaction between Earth and ocean tides. Bull. Inf. Marées Terrestres, 112, 8131–8144

Francis, O. and Mazzega, P. (1990) Global charts of ocean tide loading effects. J. Geophys. Res., 95, 11411–11424

Francis, O. and van Dam, T. (2002) Evaluation of the precision of using absolute gravimeters to calibrate superconducting gravimeters. Metrologia, 39, 485–488

Gilbert, F. and Dziewonski, A.M. (1975) An application of normal mode theory to the retrieval of structured parameters and source mechanism from seismic spectra. Phil. Trans. R. Soc. Lond. A, 278, 187–269

Goodkind, J.M. (1999) The superconducting gravimeter. Rev. Sci. Instrum., 70(11), 4131–4152

Gross, R.S., Fukumori, I. and Menemenlis, D. (2003) Atmospheric and oceanic excitations of the Earth's wobbles during 1980–2000. J. Geophys. Res., 108, 2370–2386

Guo, J.Y., Dierks, O., Neumeyer, J. and Shum, C.K. (2006) Weighting algorithm to stack superconducting gravimeter data for the potential detection of the Slichter modes. J. Geodyn., 41, 326–333

Harnisch, G. and Harnisch, M. (2002) Seasonal variations of hydrological influences on gravity measurements at Wettzell. Bull. Inf. Marees Terrestres, 137, 10849–10861

Harnisch, M. and Harnisch, G. (2006) Study of long term gravity variations based on data of the GGP co-operation. J. Geodyn., 41(1–3), 318–325

Harrison, P.J. and Stevens, C.F. (1976) Bayesian forecasting (with discussion). J. R. Stat. Soc. B, 38, 205–247

Hartmann, T. and Wenzel, H.G. (1995a) The HW95 tidal potential catalogue. Geophys. Res. Lett., 22(24), 3553–3556

Hartmann, T. and Wenzel, H.G. (1995b) Catalogue HW95 of the tide generating potential. Bull. Inf. Marees Terrestres, 123, 9278–9301

Hinderer, J., Amalvict, M., Franzis, O. and Mäkinen, J. (1998) On the calibration of superconducting gravimeters with the help of absolute gravity measurements. In: Ducarme, B. and Paquet, P. (eds) Proceedings of the 13th International Symposium on Earth Tides, Brussels, 1997, 557–564

Hinderer, J., Andersen, O., Lemoine, R., Crossley, D. and Boy, J.P. (2006) Seasonal changes in the European gravity field from GRACE: a comparison with superconducting gravimeters and hydrology model predictions. J. Geodyn., 41, 59–68

Hinderer, J. and Crossley, D. (2000) Time variations in gravity and interferences on the Earth's structure and dynamics. Surv. Geophys., 21, 1–45

Hinderer, J. and Crossley, D. (2004) Scientific achievements from the first phase (1997–2003) of the global geodynamics project using a worldwide network of superconducting gravimeters. J. Geodyn., 38(3–5), 237–262

Hinderer, J., Crossley, D. and Warburton, R. (2007) Gravimetric methods – superconducting gravity meters. Treatise Geophys., 3, 65–122

Hinderer, J. and Legros, H. (1989) Elasto-gravitational deformation, relative gravity changes and Earth dynamics. Geophys. J. Int., 97, 481–495

Hinderer, J., Legros, H. and Crossley, D. (1991) Global Earth dynamics and induced gravity changes. J. Geophys. Res., 96(B12), 20257–20265

Huang, J., Van den Dool, H.M. and Georgakakos, K.P. (1996) Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts. J. Clim., 9, 1350–1362

Imanishi, Y. (2005) On the possible cause of long period instrumental noise (parasitic mode) of a superconducting gravimeter. J. Geod., 78, 683–690

Imanishi, Y., Sato, T., Higashi, T., Sun, W. and Okubo, S. (2004) A network of superconducting gravimeters detects submicrogal coseismic gravity changes. Science, 306, 476–478, doi: 10.1126/science1101875

Jekeli, C. (1981) Alternative Methods to Smooth the Earth's Gravity Field. Department of Geodetic Science and Surveying, Ohio State University, Columbus, p. 48

Khan, S.A. and Hoyer, J. (2004) Shallow-water tides in Japan from superconducting gravimetry. J. Geod., 78, 245–150

Kim, J.W., Neumeyer, J., Kim, T.H., Woo, I., Park, H.J., Jeon, J.S. and Kim, K.D. (2009) Analysis of superconducting gravimeter measurements at MunGyung Station, Korea. J. Geodyn., 47, 180–190

Klügel, T. and Wziontek, H. (2009) Correcting gravimeters and tiltmeters for atmospheric mass attraction using operational weather models. J. Geodyn., 48, 204–210

Kroner, C. (1997) Reduktion von Luftdruckeffekten in zeitabhängigen Schwerebeobachtungen. Dissertation, Technische Universität Clausthal, Clausthal-Zellerfeld

Kroner, C. (2001) Hydrological effects on gravity data of the geodynamic observatory Moxa. J. Geodyn. Soc. Jpn., 47(1), 353–358

Kroner, C., Dierks, O., Neumeyer, J. and Wilmes, H. (2005) Analysis of observations with dual sensor superconducting gravimeters. Phys. Earth Planet. Int., 153, 210–219

Kroner, C. and Jahr, T. (2006) Hydrological experiments around the superconducting gravimeter at Moxa observatory. J. Geodyn., 41, 268–275

Kroner, C. and Jentzsch, G. (1998) Methods of air pressure reduction tested on Potsdam station. Bull. Inf. Marees Terrestres, 127, 9834–9842

Lambert, A. and Bower, D. (1991) Constraints on the usefulness of gravimetry for detecting precursory crustal deformations. Tectonophysics, 193(4), 369–375

Lambert, A., Courtier, N. and James, T.S. (2006) Long-term monitoring by absolute gravimetry: tides to postglacial rebounds. J. Geodyn., 41, 307–317

Lefevre, F., Lyard, F.H., Le Provost, C. and Schrama, E.J.O. (2002) FES99: a global tide finite element solution assimilating tide gauge and altimetry information. J. Atmos. Oceanic Technol., 19, 1345–1356

Lei, X. and Xu, H. (2002) Tri-frequency spectrum method and results for resolving the parameters of the Earth core free nutation. Sci. China Ser D, 45(4), 325–336

Lei, X., Xu, H. and Sun, H.P. (2005) Detection of spheroid free oscillation excited by Peru 8.2 Ms earthquake with five international superconducting gravimeter data. Sci. China Ser D Earth Sci., 48(1), 123–133

Le Provost, C., Lyard, F., Lefevre, F. and Roblou, L. (2002) FES 2002 – a new version of the FES tidal solution series. Abstract Volume Jason-1 Science Working Team Meeting, Biarritz

Mathews, P.M. (2001) Love numbers and gravimetric factor for diurnal tides. Proceedings of the 14th International Symposium on the Earth Tides. J. Geod. Soc. Jpn., 47(1), 231–236

Mathews, P.M., Herring, T.A. and Buffett, B.A. (2002) Modelling of nutation precision: new nutation series for nonrigid Earth and inside into the Earth's interior. J. Geophys. Res., 107(B4), ETG3-1, 3–30

Matsumoto, K., Takanezawa, T. and Ooe, M. (2000) Ocean tide models developed by assimilating TOPEX/POSEIDON altimeter data into hydrodynamic model, a global model and regional model around Japan. J. Oceanogr., 56, 567–581

Matsumoto, K.T., Sato, T., Takanezawa, T., Ooe, M. (2001) GOTIC2: a program for computation of oceanic tidal loading effect. J. Geod. Soc. Jpn., 47, 243–248

McCarthy, D. and Petit, G. (2004) IERS Conventions (2003). IERS Technical Note 32, IERS Central Bureau, Frankfurt am Main.

Melchior, P., Moens, M., Ducarme, B. and van Ruymbeke, M. (1981) Tidal loading along a profile Europe–East Africa–South Asia–Australia and the Pacific Ocean. Phys. Earth Planet. Int., 25, 71–106

Merriam, J.B. (1992) Atmospheric pressure and gravity. Geophys. J. Int., 109, 488–500

Meurers, B. (1999) Air pressure signatures in the SG data of Vienna. Bull. Inf. Marees Terrestres, 131, 10195–10200

Meurers, B., Van Camp, M. and Petermans, T. (2007) Correcting superconducting gravity time-series using rainfall modelling at Vienna and Membach station and application to Earth tide analysis. J. Geod., 81(11), 703–712

Meurers, B., Van Camp, M., Petermans, T., Verbeeck, K. and Vanneste, K. (2005) Investigation of local atmospheric and hydrological gravity signals in superconducting gravimeter time series. Geophys. Res. Abstr., 7, 07463

Milly, P.C.D. and Shmakin, A.B. (2002) Global modeling of land water and energy balances. Part I: The Land Dynamics (LaD) model. J. Hydrometeorol., 3(3), 283–299

Nawa, K., Suda, N., Fukao, Y., Sato, T., Tamura, Y., Shibuya, K., McQueen, H., Virtanen, H. and Kääriäinen, J. (2000) Incessant excitation of the Earth's free oscillations: global comparison of superconducting gravimeter records. Phys. Earth Planet. Int., 120, 289–297

Neuberg, J. (1987) Erdkernresonanz im ganztägigen Gezeitenband. Dissertation, Universitat Karlsruhe, Karlsruhe

Neuberg, J., Hinderer, J. and Zürn, W. (1987) Stacking gravity tide observation in central Europe for the retrieval of the complex eigenfrequency of the nearly diurnal free wobble. Geophys. J. Int., 91, 853–868

Neumeyer, J. (1995) Frequency dependent atmospheric pressure correction on gravity variations by means of cross spectral analysis. Bull. Inf. Marees Terrestres, 122, 9212–9220

Neumeyer, J., Barthelmes, F., Combrinck, L., Dierks, O. and Fourie, P. (2002) Analysis results from the SG registration with the Dual Sphere Superconducting Gravimeter at SAGOS (South Africa). Bull. Inf. Marees Terrestres, 135, 10607–10616

Neumeyer, J., Barthelmes, F., Dierks, O., Flechtner, F., Harnisch, M., Harnisch, G., Hinderer, J., Imanishi, Y., Kroner, C., Meurers, B., Petrovic, S., Reigber, Ch., Schmidt, R., Sun, H.P. and

Virtanen, H. (2006) Combination of temporal gravity variations resulting from superconducting gravimeter (SG) recordings, GRACE satellite observations and hydrology models. J. Geod., 79, 573–585

Neumeyer, J., Barthelmes, F., Kroner, C., Petrovic, S., Schmidt, R., Virtanen, H. and Wilmes, H. (2008) Analysis of gravity field variations derived from Superconducting Gravimeter recordings, the GRACE satellite and hydrological models at selected European sites. Earth Planet Space, 60, 505–518

Neumeyer, J., Barthelmes, F. and Wolf, D. (1998) Atmospheric pressure correction for gravity data using different methods. Proceedings of the 13th International Symposium on Earth Tides, Brussels, 1997

Neumeyer, J., del Pino, J., Dierks, O. and Pflug, H. (2005) Improvement of ocean loading correction on gravity data with additional tide gauge measurements. J. Geodyn., 40, 104–111

Neumeyer, J., Hagedoorn, J., Leitloff, J. and Schmidt, T. (2004) Gravity reduction with three-dimensional atmospheric pressure data for precise ground gravity measurements. J. Geodyn., 38(3–5), 437–450

Neumeyer, J., Schmidt, T. and Stoeber, C. (2007) Improved determination of the atmospheric attraction with 3D air density data and its reduction on ground gravity measurements. International Association of Geodesy Symposia, Dynamic Planet, Cairns, Australia, Vol. 130. Springer, Berlin, Heidelberg, New York, pp. 541–548

Niebauer, T., Sasagawa, G., Faller, J., Hilt, R., Klopping, F. (1995) A new generation of absolute gravimeters. Metrologia, 32, 159–180

Penna, N.T., Bos, M.S., Baker, T.F. and Scherneck, H.G. (2008) Assessing the accuracy of predicted ocean tide loading displacement values. J. Geod., doi: 10.1007/s00190-008-0220-2

Peterson, J. (1993) Observation and modelling of background seismic noise. Open File Report 93-332, US Department of Interior, Geological Survey, Albuquerque

Pick, M., Picha, J. and Vyskocil, V. (1973) Theory of the Earth`s Gravity Field. Publishing House of the Czechoslovak Academy of Sciences, Prague

Prothero, W.A. and Goodkind, J.M. (1968) A superconducting gravimeter. Rev. Sci. Instrum., 39(9), 1257–1262

Rehen, F. (1997) Relative gravity measurements with a Scintrex CG-3 M in the gravimeter calibration systems Hannover and Hornisgrinde. Bur. Grav. Int. Bull. d'Inf. Toulouse, 81, 23–29

Reigber, Ch., Schmidt, R., Flechtner, F., König, R., Meyer, U., Neumayer, K.H., Schwintzer, P. and Zhu, S.Y. (2005) An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S. J. Geodyn., 39, 1–10

Richter, B. and Warburton, R. (1989) A new generation of superconducting gravimeters. In: Ducarme, B. and Paquet, E (eds) Proceedings of the 13th International Symposium on Earth Tides, Brussels, 1997, pp. 545–555

Richter, B. and Wenzel, H.G. (1991) Precise instrumental phase lag determination by the step response method. Bull. Inf. Marees Terrestres, 111, 8032–8052

Rogers, G. and Dragert, H. (2003) Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip. Science, 300, 1942–1943

Rogister, Y. (2003) Splitting of seismic free oscillations and the Slichter triplet using normal mode theory of a rotating ellipsoidal Earth. Phys. Earth Planet. Int., 148, 169–182

Rosat, S., Hinderer, J., Crossley, D. and Boy, J.P. (2004) Performance of superconducting gravimeters from long-period seismology to tides. J. Geodyn., 38(3–5), 461–476

Rosat, S., Rogister, Y., Crossley, D. and Hinderer, J. (2006). A search for the Slichter triplet with superconducting gravimeters: impact on the density jump at the inner core boundary. J. Geodyn., 41, 296–306

Rosat, S., Sato, T. and Imanishi, Y. (2005) High resolution analysis of the gravest seimic normal modes after the 2004 M = 9 Sumatra earthquake using superconducting gravimeter data. Geophys. Res. Lett., 32, L13304

Sato, T., Tamura, Y., Matsumoto, K., Imanishi, Y. and Mac Queen, H. (2004) Parameters of the fluid core resonance inferred from superconducting gravimeter data. J. Geodyn., 38, 375–389

Schmidt, R., Flechtner, F., Meyer, Ul., Reigber, Ch., Barthelmes, F., Foerste, Ch., Stubenvoll, R., König, R., Neumayer, K.H. and Zhu, S.Y. (2005) Static and time-variable gravity from GRACE mission data. In: Flury, J., Rummel, R., Reigber, C., Rothacher, M., Boedecker, G. and Schreiber, U. (eds) Observation of the Earth System from Space. Springer, Berlin, Heidelberg, New York, pp. 115–129, 293–296

Schmidt, R., Schwintzer, P., Flechtner, F., Reigber, Ch., Güntner, A., Döll, P., Ramillien, G., Cazenave, A., Petrovic, S., Jochmann, H. and Wünsch, J. (2006) GRACE observations of changes in continental water storage. Glob. Planet. Change, 48(4), 259–273

Schüller, K. (1976) Ein Beitrag zur Auswertung von Erdgezeitenregistrierungen, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Nr. 227, München

Simon, D. (2002) Modelling of the field of gravity variations induced by the sectional air masses warming during 1998–2000. Bull. Inf. Marees Terrestres, 136, 10821–10838

Slichter, L.B. (1961) The fundamental free mode of the Earth's inner core. Proc. Natl. Acad. Sci. USA, 47, 186–190

Smylie, D.E., Jiang, X., Brennan, B.J. and Sata, K. (1992) Numerical calculation of the modes of oscillation of the Earth's core. Geophys. J. Int., 108, 465–490

Smylie, D., Francis, O. and Merriam, J. (2001) Beyond tides – determination of the core properties from superconducting gravimeter observations. J. Geod. Soc. Jpn., 47(1), 364–372

Smylie, D.E., Hinderer, J., Richter, B. and Ducarme, B. (1993) The product spectra of gravity and barometric pressure in Europe. Phys. Earth Planet. Int., 80, 135–157

Sun, H.P. (1995) Static deformation and gravity changes at the Earth′s surface due to the atmospheric pressure. Observatoire Royal des Belgique, Serie Geophysique Hors-Serie, Bruxelles

Sun, H.P., Hsu, H., Chen, W., Chen, X.D., Zhou, J.C., Liu, M. and Gao, S. (2006) Study of Earth's gravity tide and ocean loading characteristics in Hongkong area. Chinese J. Geophys., 49(3), 657–670

Sun, H.P., Hsu, H., Luo, S. and Xu, J. (1999) Study of the ocean models using tidal gravity observations obtained with superconducting gravimeter. Acta Geodetica et Cartograophy Sinica, 28(2), 64–71

Sun, H.P., Hsu, H.T., Jentzsch, G. and Xu, J.Q. (2002a) Tidal gravity observations obtained with superconducting gravimeter and its application to geodynamics at Wuhan/China. J. Geodyn., 33(1–2), 187–198

Sun, H.P., Jiangiao, X.V. and Ducarme, B. (2003) Experimental Earth tide models in considering free wobble of the Earth's liquid core. Chin. Sci. Bull., 48(4), 935–940

Sun, H.P., Xu, J. and Ducarme, B. (2004) Detection of the translational oscillation of the Earth's solid inner core based on the international superconducting gravimeter observations. Chinese Sci. Bull., 49(11), 1165–1178

Sun, H.P., Xu, J.Q. and Decurme, B. (2002b) Experimental Earth tidal models of the core resonance obtained by stacking tidal gravity observations from GGP stations. Bull. Inf. Marees Terrestres, 136, 10729–10730

Tamura, Y. (1987) A harmonic development of tide-generating potential. Bull. Inf. Marees Terrestres, 99, 6813–6855

Tamura, Y. (1990) BAYTYP-G Users Manual. National Astronomical Observatory, Mizusawa

Tamura, Y., Sato, T., Ooe, M. and Ishiguro, M. (1991) A procedure for tidal analysis with a Bayesian information criterion. Geophys. J. Int., 104, 507–516

Tanaka, Y., Okubo, S., Machida, M., Kimura, I. and Kosuge, T. (2001) First detection of absolute gravity change caused by earthquake. Geophys. Res. Lett., 28(15), 2979–2981

Tapley, B.D. and Reigber, Ch. (2001) The GRACE mission: status and future plans. EOS Trans. Am. Geophys. Union, 82(47), Fall Meet. Suppl., G41 C-02

Timmen, K. and Gitlein, O. (2004) The capacity of the Scintrex Autograv CG-3 M no. 4492 gravimeter for "absolute-scale" surveys. In: de Menezes P.M.L. (ed) Revista Brasileira de Cartografia (Brazilian Journal of Cartography), 56/02, 89–95

Torge, W. (1989) Gravimetry. de Gruyter, Berlin, New York

Van Camp, M. (1999) Measuring normal seismic modes with the GWR C021 superconducting gravimeter. Phys. Earth Planet. Int., 116, 81–92

Van Camp, M. and Francis, O. (2007) Is the drift of superconducting gravimeters a linear or exponential function of time? J. Geod., 81, 335–344

Van Camp, M. and Vauterin, P. (2005) Tsoft: graphical and interactive software for the analysis of time series and Earth tides. Comput. Geosci., 31(5), 631–640

Van Camp, M., Wenzel, H.G., Schott, P., Vauterin, P. and Francis, O. (2000) Accurate transfer function determination for superconducting gravimeters. Geophys. Res. Lett., 27(1), 37–40

Van Dam, T. and Francis, O. (1998) Two years of continuous measurements of tidal and non tidal variations in gravity in Boulder, Colorado. Geophys. Res. Lett., 25(3), 393–396

Vanicek, P. and Krakiwsky, E.J. (1982) Geodesy: The Concepts. North-Holland Publishing Company, Amsterdam, New York

Vauterin, P. (1998) The correction of the pressure effect for the Superconducting Gravimeter in Membach (Belgium). In: Ducarme, B. and Paquet, P. (eds) Proceedings of the 13th International Symposium on Earth Tides, Brussels, 1997, pp. 447–454

Venedikov, A.P. (1966a) Une méthode d'analyse des Marées terrestres à partir d'enregistrements des longueur arbitraires. Bull. Cl. Sci., 5^{e} S., t. LIII, fasc 3, Communications de l'Observatoire Royal de Belgique, Série géoph., 71, 463–485

Venedikov, A.P. (1966b) Sur la constitution de filters numériques pour le traitement des enregistrements des Marées terrestres. Bull. Cl. Sci., 5^{e} S., t. LIII, fasc 6, Communications de l'Observatoire Royal de Belgique, Série géoph., 76, 827–845

Venedikov, A.P., Arnoso, J. and Vieira, R. (2001) Program VAV/2000 for tidal analysis of unevenly spaced data with irregular drift and coloured noise. J. Geod. Soc. Jpn., 47(1), 281–286

Venedikov, A.P., Vieira, R., de Toro, C. and Arnoso, J. (1997) A new program developed in Madrid for tidal data processing. Bull. Inf. Marees Terrestres, 126, 9669–9704

Virtanen, H. (1996) Observations of free oscillations of the Earth by superconducting gravimeter GWT T020. Acta Geod. Geophys. Hung., 33, 429–431

Virtanen, H. (2001) Hydrological studies at the gravity station Metshovi, Finland. J. Geod. Soc. Jpn., 47(1), 328–333

Virtanen, H. (2006) Studies of Earth Dynamics with the Superconducting Gravimeter. Academic Dissertation in Geophysics, University of Helsinki, Helsinki, http://ethesis.helsinki.fi/

Virtanen, H., Bilker, M., Mäkinen, J., Poutanen, M., Tervo, M., Vehviläinen, B., Huttunen, M. and Mäkinen, R. (2005) Comparison of modelled variation in water storage in Finland with GRACE and superconducting gravimeter observations. Geophys. Res. Abst., 7, 06248

Wahr, J. (1985) Deformation induced by polar motion. J. Geophys. Res., 90(B11), 9363–9368

Wahr, J., Molenaar, M. and Bryan, F. (1998) Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. J. Geophys. Res., 103(12), 30205–30229

Wahr, J., Swenson, S., Zlotnicki, V. and Velicogna, I. (2004) Time-variable gravity from GRACE: first results. Geophys. Res. Lett., 31(11), L11501, doi: 10.1029/2004GL019779

Warburton, R.J. and Brinton, E. (1995) Recent developments in GWR Instruments superconducting gravimeters. Proceedings of the 2nd Workshop: Non-tidal Gravity Changes. Intercomparison Between Absolute and Superconducting Gravimeters, Vol. 11. Cahiers du Centre Europeen de Geodynamique et de Seismologie, Luxembourg, pp. 23–56

Warburton, R.J. and Goodkind, J.M. (1977) The influence of barometric – pressure variations on gravity. Geophys. J. R. Astr. Soc., 48, 281–292

Warburton, R.J., Brinton, E., Reineman, R. and Richter, B. (2000) Remote operation of superconducting gravimeters. Proceedings of the Workshop: High Precision Gravity Measurements with Applications to Geodynamics and Second GGP Workshop, Vol. 17. Cahiers du Centre Europeen de Geodynamique et de Seismologie, Luxembourg, pp. 125–136

Wenzel, H.G. (1995a) Accurate instrumental phase lag determination for feedback gravimeters. In: Hsu H.T. (ed) Proceedings 12th International Symposium on Earth Tides. Science Press, Beijing, New York, pp. 191–198

Wenzel, H.G. (1995b) Gezeitenpotential. In: Bericht zum DGG-Seminar Gezeiten in Oberwolfach/Schwarzwald Oktober 1994, Sonderband II, Deutsche Geophysikalische Gesellschaft e.V. Münster, pp. 1–18

Wenzel, H.G. (1995c) Erdgezeitenanalyse. In: Bericht zum DGG-Seminar Gezeiten in Oberwolfach/Schwarzwald Oktober 1994, Sonderband II, Deutsche Geophysikalische Gesellschaft e.V. Münster, pp. 19–38

Wenzel, H.G. (1996) The nanogal software: Earth tide data processing package ETERNA 3.30. Bull. Inf. Marees Terrestres, 124, 9425–9439

Widmer-Schmidrig, R. (2003) What can superconducting gravimeters contribute to normal mode seismology? Bull. Seism. Soc. Am., 93(3), 1370–1380

Wilson, C.R., Wu, H., Scanlon, B. and Sharp, J.M. (2007) Taking the superconducting gravimeter to the field for hydrologic and other investigations. EOS Trans. Am. Geophys. Union, Fall Meeting 2007, paper H11A-050

Xu, G. (2003/2007) GPS – Theory, Algorithms and Applications, 1st/2nd Eds. Springer, Berlin

Xu, J., Sun, H.P. and Luo, S. (2002) Study of the Earth`s free core nutation by tidal gravity data recorded with international superconducting gravimeters. Sci. China Ser D, 45(4), 337–347

Xu, J.Q., Sun, H.P. and Yang, X.F. (2004) A study of gravity variations caused by polar motion using superconducting gravimeter data from the GGP network. J. Geod., 78, 201–209

Zerbini, S., Richter, B., Negusini, M., Romagnoli, C., Simon, D., Domenichini, F. and Schwahn, W. (2001) Height and gravity variations by continuous GPS, gravity and environmental parameter observations in the southern Po Plain, near Bologna, Italy. Earth Planet. Sci. Lett., 192, 267–279

Zürn, W. (2002) Simplistic models of vertical seismic noise above 0.1 mHz derived from local barometric pressure. Bull. Inf. Marees Terrestres, 137, 10876–10874

Zürn, W. and Wielandt, E. (2007) On the minimum of vertical seismic noise near 3 mHz. Geophys. J. Int., 168, 647–658

Zürn, W. and Wilhelm, H. (1984) Tides of the Earth, Vol. 2. In Landolt-Börnstein, Springer, Berlin, Heidelberg, New York, Tokyo, pp. 259–299

Chapter 11
Synthetic Aperture Radar Interferometry

Ye Xia

Contents

Y. Xia (✉)
German Research Centre for Geosciences, Telegrafenberg A17, Potsdam D-14473, Germany
e-mail: xia@gfz-potsdam.de

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1_11,

11.1 Introduction

In the past two decades INSAR technique (synthetic aperture radar interferometry) has been quickly developed and widely used for the study of topography (digital elevation model generation) and deformation (Earth surface monitoring). Synthetic aperture radar (SAR) is a coherent active microwave image instrument, which is used for mapping the scattering properties of the Earth's surface in the respective wavelength domain. The intensity (grey value) of each pixel in an SAR image represents the physical and geological property and geometric parameters of the imaged scene. In the late 1970s the first space-borne SAR system in the world (NASA satellite SEASAT) was launched for Earth observation. This mission pioneered the application of the SAR technology to mapping the Earth's surface, acquiring information about physical and geological properties such as topography, morphology, moisture, and finding underground water. The space-borne SAR systems operate in the microwave range and therefore work on all days and all meteorological conditions.

Compared to optical imagery, besides intensity the SAR images provide the second physical value for each of their pixel: phase, which represents the time delay of the radar pulse from the illuminated object. The phase difference of two SAR images acquired in same scene, but at different sensor positions, may offer the possibility to extract the topography and detect the displacement of the acquired surface. Based on the concept, the principle and applications of the SAR interferometry technology became popular recently. In the 1980s the first results were published (Goldstein and Zebker 1987; Goldstein et al. 1988). Since 1991, after the launch of the ESA satellite ERS-1, followed by ERS-2 (1995), large numbers of SAR data sets have been offered to investigators for their scientific research and applications. Then more and more publications about the SAR interferometry method and applications have been published. Now SAR interferometry becomes an extremely powerful tool for mapping the Earth's surface topography and detecting associated movement over a large spatial scale with centimetre, and even millimetre, accuracy. Today differential SAR interferometry (D-INSAR) is one of the most important methods for earthquake and volcanic deformation research, for glaciology and ice sheet monitoring, for monitoring of land surface subsidence due to mining, gas or groundwater extraction, landslide and other geologic hazards.

The purpose of this chapter is to introduce the principles and data processing of SAR interferometry including differential SAR interferometry, corner reflector SAR interferometry (CR-INSAR) and some practical applications. First, in Sect. 11.2 the basics of SAR imaging are briefly reviewed to understand the SAR imaging process and SAR image feature, which is also the background of the SAR interferometry. Section 11.3 describes the principle and data processing of SAR interferometry for digital elevation model (DEM) generation. Section 11.4 deals with differential SAR interferometry. In this section differential interferometry of the persistent coherent scatterers is mainly introduced for monitoring of the subsidence in the area covered by vegetables.

11.2 Synthetic Aperture Radar Imaging

Imaging radar is classified into real aperture radar (RAR) and synthetic aperture radar (SAR). RAR transmits a narrow angle beam of microwave pulse in the range direction at right angles (side-looking) to the flight direction (azimuth direction) and receives the backscattering from the targets which will be transformed to a radar image from the received signals. Usually the reflected pulse will be arranged in the order of return time from the targets, which corresponds to the range direction scanning. Figure 11.1 shows the RAR illumination geometry. The resolution in the range direction ρ_r depends on the pulse width. However, if the pulse width is made small, in order to increase the resolution, the S/N ratio of the return pulse will decrease because the transmitted power also becomes low. Therefore, the transmitted pulse is modulated to chirp with a high power but wide band, which is received through a matched filter, with reverse function of transmission, to make the pulse width very narrow and high power. The resolution in the azimuth direction ρ_a is equal to the radar aperture L_s, i.e. multiplication of the beam width β and the distance to a target R shown in Fig. 11.2. The beam width is determined by the ratio of the wavelength λ and the antenna length D. Therefore, the resolution in the azimuth direction ρ_a increases with shorter wavelength and bigger antenna size:

$$\rho_a = L_s = \frac{\lambda}{D} R, \tag{1}$$

where λ is the wavelength, D is the radar antenna length in flight direction and R is the distance of the target away from radar. However, it is difficult to attach a large antenna. For example, in order to obtain 5-m resolution in azimuth direction with C band ($\lambda = 5$ cm) and 850 km distance from a target, an antenna with 8.5-km diameter is needed. The real aperture radar therefore has a technical limitation for practical application.

Compared to real aperture radar, synthetic aperture radar (SAR) synthetically increases the antenna's size or aperture to increase the azimuth resolution through

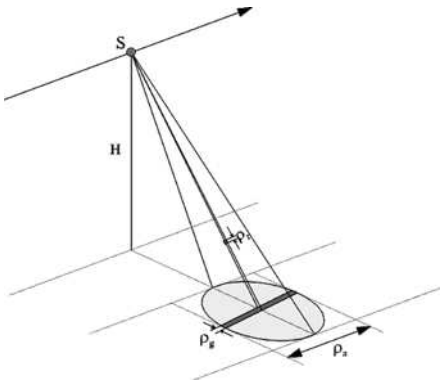


Fig. 11.1 Real aperture radar geometry

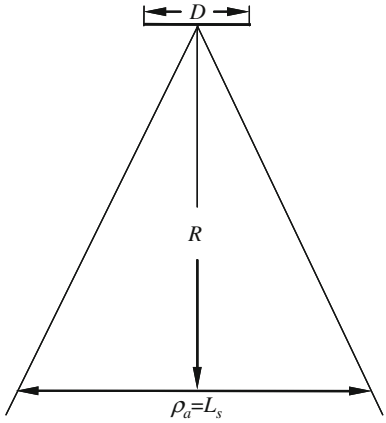


Fig. 11.2 Aperture of the RAR

the same pulse compression technique as explained for range direction. In this section we will briefly discus how to reconstruct an SAR image from raw data and how to increase synthetically the antenna's size (aperture) to improve the azimuth resolution, as well as how to understand the SAR image features.

Like other remote sensing systems, an imaging radar sensor may be carried on either an airborne or a space-borne platform. As with any aircraft, the airborne radar will be susceptible to variations in velocity and other motions of the aircraft as well as to environmental (weather) conditions. In order to avoid image artefacts or geometric positioning errors due to random variations in the motion of the aircraft, the radar system must use sophisticated navigation/positioning equipment. The space-borne radars are not affected by motion of this type, and the geometry of their orbits is usually very stable and their positions can be accurately calculated. In this and following sections we discus about space-borne SAR only and the main radar parameters are from ESA ENVISAT ASAR.

11.2.1 Radar Transmitted and Received Signal

A space-borne SAR, having the viewing geometry shown in Fig. 11.3, consists of a sensor path and an illuminated surface (Xia 1996). The radar transmitted signal is a swept linear frequency modulated (FM) pulse, which can be expressed as

$$s(t) = \mathrm{rect}\left(\frac{t}{\tau_p}\right) \times \mathrm{e}^{j2\pi\left(f_0 t + \frac{1}{2}kt^2\right)}, \tag{2}$$

where rect() is a rectangular window function being unity over the pulse duration τ_p and zero elsewhere, which is also called range pulse window, and f_0 is the radar carrier frequency, which is modulated with a frequency rate of k. Typical SAR currier wavelength (light speed/f_0) are approximately 3 cm (X band), 6 cm (C band), 9 cm

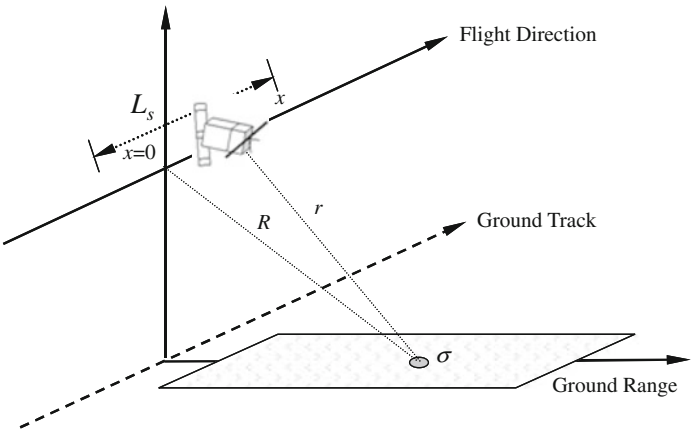


Fig. 11.3 Simplified SAR geometry

(S band) and 24 cm (L band). The band width B of the transmitted FM is $k\tau$, which significantly improves the range resolution of the SAR.

Assume that a point target with a backscatter coefficient σ is located on the Earth's surface; the shortest distance between the radar along-track and the point target is R, so the echo signal from the point target is

$$s_e(t) = \mathrm{rect}\left(\frac{t-\tau}{\tau_\mathrm{p}}\right) \times \mathrm{e}^{j2\pi f_0(t-\tau)} \times \mathrm{e}^{j\pi k(t-\tau)^2}, \tag{3}$$

where τ is the delay of the radar signal and depends on the range (radar–target distance) r:

$$\begin{aligned} \tau &= \tfrac{2r}{c} \\ &\approx \tfrac{2R}{c}\left(1+\tfrac{x^2}{2R^2}\right), \end{aligned} \tag{4}$$

where x is the location of the radar on its along-track and is equal to vt, and v is the radar's speed. Because the radar transmits a train of FM pulse with a period of 1/PRF, PRF is the pulse repetition frequency. Usually the period is about hundreds of microseconds; therefore, the radar–target distance variation r can be seen as a slow-varying function over the pulse period relative to change of the radar transmitted waveform. Hence, the echo signal can be expressed as a two-dimensional signal:

$$s_e(x,t) = \mathrm{rect}\left(\frac{t-\dfrac{2R}{c}}{\tau_\mathrm{p}}\right) \times \mathrm{e}^{j\left(2\pi f_0 t - \frac{4\pi R}{\lambda}\right)} \times \mathrm{e}^{-j\frac{2\pi}{\lambda R}x^2} \times \mathrm{e}^{j\pi k\left(t-\frac{2R}{c}\right)^2}. \tag{5}$$

When we only consider this case, in which the shortest distance between radar and target is R, after demodulation the echo train $s_e(x,t)$ received by the radar can be expressed as

$$s_e(x,t:R) = \text{rect}\left(\frac{t}{\tau_p}\right) \times \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \mathrm{e}^{j\pi k\left(t-\frac{2R}{c}\right)^2} \times \mathrm{e}^{-j\frac{2\pi}{\lambda R}x^2}. \tag{6}$$

Equation (6) represents a point target response of the synthetic aperture radar with a parameter R being the shortest range distance between sensor and the point target and from which the unit impulse response of the SAR can be derived.

11.2.2 Impulse Response of SAR

According to the linear system theory, the impulse response of a linear system is the autocorrelation function of input signal of the linear system (optimal filter). Suppose SAR is a linear system, whose impulse response can be represented as follows:

$$\begin{aligned} h(x,t;R) &= \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \int_{-\frac{\tau_p}{2}}^{\frac{\tau_p}{2}} \int_{-\frac{L_s}{2}}^{\frac{L_s}{2}} \left[\mathrm{e}^{j\pi k(t')^2} \times \mathrm{e}^{-j\frac{2\pi}{\lambda R}x'^2}\right] \times \left[\mathrm{e}^{-j\pi k(t'+t)^2} \times \mathrm{e}^{j\frac{2\pi}{\lambda R}(x'+x)^2}\right] \times dt'\, dx' \\ &= \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \int_{-\frac{\tau_p}{2}}^{\frac{\tau_p}{2}} \left[\mathrm{e}^{j\pi k(t')^2} \times \mathrm{e}^{-j\pi k(t'+t)^2}\right] \times dt' \int_{-\frac{L_s}{2}}^{\frac{L_s}{2}} \left[\mathrm{e}^{-j\frac{2\pi}{\lambda R}x'^2} \times \mathrm{e}^{j\frac{2\pi}{\lambda R}(x'+x)^2}\right] \times dx' \\ &= \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \tau_p\, \mathrm{e}^{-j\pi k t^2} \frac{\sin(\pi k \tau_p t)}{\pi k \tau_p t} \times L_s\, \mathrm{e}^{j\frac{2\pi}{\lambda R}x^2} \frac{\sin\left(\frac{2\pi}{\lambda R}L_s x\right)}{\frac{2\pi}{\lambda R}L_s x} \\ &= \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \tau_p\, \mathrm{e}^{-j\pi k t^2} \frac{\sin\left(\pi \frac{t}{\rho_r}\right)}{\pi \frac{t}{\rho_r}} \times L_s\, \mathrm{e}^{j\frac{2\pi}{\lambda R}x^2} \frac{\sin\left(\pi \frac{x}{\rho_a}\right)}{\pi \frac{x}{\rho_a}} \end{aligned} \tag{7}$$

where range resolution ρ_r=$1/k\tau = 1/B$, depending on the band width of the radar transmitted FM signal, and azimuth resolution $\rho_a = \lambda R/2L_s = D/2$, L_s is the aperture of the antenna with the length D; see Fig. 11.2. Now we can compare the azimuth resolution of the real aperture radar and the synthetic aperture radar. The length of their antennas is the same, but the azimuth resolution of the SAR is half of the antenna length and independent of the radar–target distance, while that of the RAR is equal to real aperture L_s depending on the shortest radar–target distance R. The range resolution ρ_r in (7) is defined in time domain, which can be converted to spatial domain, i.e. $\rho_r = c/2B$; here c is the speed of light, by ERS satellite, B is 15 MHz, so that the slant range resolution is about 10 m; the corresponding ground range resolution is about 25 m.

The impulse response (7) can be rewritten as

$$h(x,t:R) = h_r(x,t:R) \times h_a(x,t:R), \tag{8}$$

where τ_p and L_s are neglected, and

$$\begin{cases} h_r(x,t:R) = \mathrm{e}^{-j\frac{4\pi R}{\lambda}} \times \sin c\left(\frac{\pi t}{\rho_r}\right) \\ h_a(x,t:R) = \mathrm{e}^{j\frac{2\pi}{\lambda R}x^2} \sin c\left(\frac{\pi x}{\rho_a}\right) \end{cases} \tag{9}$$

The SAR's impulse response then consists of two terms with two separable variables x and t: impulse response in the range and azimuth direction, respectively. Note that the radar–target distance r is a function of t, the target field can be expressed as $\sigma(x, r)$ and radar impulse responses can be given by $h_r(x, r)$ and $h_a(x,r)$. Therefore it is clear that the targets $\sigma(x, r)$ can be reconstructed by sequentially correlating the return signal by the responses $h_r(x, r)$ and $h_a(x,r)$. We can see that if a point target has a shortest radar–target distance R, the output of the SAR or, say, the pixel value of this point target in the SAR image will include the phase $-4\pi R/\lambda$, which is a function of the SAR geometry and the topography of the point target. It is precisely due to the phase information of the SAR image that the SAR interferometry technique has been quickly developed and widely used in many scientific and practical engineering areas.

11.2.3 Pulse Compression (Focus) and Doppler Frequency

Equation (9) is a mathematic model of the SAR system only. In fact, an SAR image is produced by convolving the return signal $s_e(x, t: R)$ (6), first with the radar transmitted chirp (FM signal), called range compression or range focus, and then with an azimuth matched filter called azimuth compression. This is similar to those of range compression. The output of the range compression is also an FM signal, called Doppler signal, but an exception is that the FM rate and the centred frequency vary with the range distance. The pulse compression process is also called focus.

11.2.3.1 Range Compression

As given above, in the simplest form, the received demodulated radar signal from a single point target can be generally expressed as

$$s_e(t) = \text{rect}\left(\frac{t}{\tau_p}\right) \times \mathrm{e}^{-j\frac{4\pi r(t)}{\lambda}} \times \mathrm{e}^{j\pi k t^2}. \tag{10}$$

Based on the optimal filter theory, the matched filter for the range compression is

$$\begin{aligned} h(t) &= s^*(-t) \\ &= \text{rect}\left(\tfrac{t}{\tau_p}\right) \times \mathrm{e}^{-j\pi k t^2}. \end{aligned} \tag{11}$$

After convolving the return signal by the matched filter, assume that the radar–target distance $r(t)$ is constant during the pulse duration, and neglect the amplitude being a constant related to the chirp duration τ_p and FM rate k, the output signal of the matched filter is approximately given by

$$s_a(t) = \mathrm{e}^{-j\frac{4\pi r(t)}{\lambda}} \times \sin c(\pi k \tau_p t). \tag{12}$$

The radar can capture the point target in a range of L_s, in which the radar–target distance varies with the azimuth time t, or say low time compared to range time. Because of flight of the satellite on its along-track, which causes variation of the radar–target distance, the azimuth signal is a two-dimensional signal, and this distance variation is called range migration.

11.2.3.2 Azimuth Compression

Concerning the azimuth signal, (11), this is a train of compressed pulses, of which envelope is a sinc function. Now consider the azimuth phase only:

$$\varphi(t) = -\frac{4\pi \left|\overrightarrow{r(t)}\right|}{\lambda}, \tag{13}$$

where $\overrightarrow{r(t)}$is the distance vector and

$$\overrightarrow{r(t)} = \overrightarrow{r(0)} + \overrightarrow{V}\, t + \frac{1}{2}\, \overrightarrow{A}\, t^2, \tag{14}$$

where $\overrightarrow{V}$and $\overrightarrow{A}$ are the velocity and acceleration vectors of the sensor, respectively. The amplitude of the distance vector is given as

$$\left|\overrightarrow{r(t)}\right| \approx \left|\overrightarrow{r(0)}\right| + \frac{\left|\overrightarrow{r(0)}\right| \times \left|\overrightarrow{V}\right|}{\left|\overrightarrow{r(0)}\right|} t + \frac{\left|\overrightarrow{r(0)}\right| \times \left|\overrightarrow{A}\right| + \left|\overrightarrow{V}\right|^2}{2\left|\overrightarrow{r(0)}\right|} t^2. \tag{15}$$

The Doppler frequency, caused by the relative motion between radar and target, is derived as

$$\begin{aligned} f_d(t) &= \frac{1}{2\pi} \times \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} \\ &= -\frac{2}{\lambda} \times \frac{\left|\overrightarrow{r(0)}\right| \times \left|\overrightarrow{V}\right|}{\left|\overrightarrow{r(0)}\right|} - \frac{2}{\lambda} \times \frac{\left|\overrightarrow{r(0)}\right| \times \left|\overrightarrow{A}\right| + \left|\overrightarrow{V}\right|^2}{2\left|\overrightarrow{r(0)}\right|} t. \\ &= f_{\mathrm{dc}} + f_{\mathrm{dr}} t \end{aligned} \tag{16}$$

Equation (16) suggests that the Doppler signal is also a frequency modulated pulse, f_{dc} is the Doppler frequency centroid and f_{dr} is the Doppler frequency rate. The azimuth compression process is similar to those of range compression, but with different frequency centres and frequency rates. In order to design the azimuth matched filter, the accurate estimation of the Doppler centroid and Doppler frequency rate is essential. It is also to be noted that in the azimuth compression process the range migration has to be corrected exactly. According to the theory of the linear system, synthetic aperture radar can be considered as a cascade-connected system consisting of two matched filters, the range matched filter and the azimuth matched filter; the

input is the raw radar data and the output is the processed SAR image. The image amplitude shows the physical backscatter characteristics of the illuminated targets, while the image phase provides the distance information between the sensor and the illuminated targets. Later it will be described how to use the SAR image phase information to measure the topography and change of the Earth's surface.

It is known that the radar can capture a target located on the Earth's surface only in a certain segment of the along-track limited by the corresponding synthetic aperture and the transmitted centre pulse is perpendicular to the flight path; this geometry and operating type is called *strip mode* of the SAR resulting in a strip map with a certain swath. The achievable azimuth resolution in the strip-mode SAR is equal to one-half of the width of the antenna. An important condition that accompanies this resolution limit, however, is that pulses have to be transmitted with spacing along flight track, which is equal to the resolution. There are two other SAR-operating modes: the spotlight mode and the scan mode, the former being for improving the azimuth resolution with fewer collected radar pulses compared to the conventional strip-mode SAR, while the latter for increasing the width of the radar scanning swath without increasing the antenna size. The principle and performance of the new modes will be briefly described as follows, because the main goal of the chapter is not for practical and detailed SAR imaging data process, but only for understanding them and then to use them for establishing the SAR interferometry foundation and performing the interferometric data process. Figure 11.4 is an ENVISAT ASAR image of Tianjin region including Tianjin, Langfang and Tanggu; the coverage area is 100 km $\times$100 km with a resolution of 30 m acquired on 2007-02-23-022552.

11.2.4 Spotlight Mode

In spotlight-mode collection, the radar antenna is continuously steered to keep the pulse beam always on the targets in a small region. Figure 11.5 shows the flight geometry of the spotlight-mode radar collection. The name for this mode comes from the fact that the radar is actually spotlighting the targets for a long time, causes a longer synthetic aperture and can therefore achieve a higher azimuth resolution.
In order to understand the spotlight-mode SAR principle for higher azimuth resolution at the expense of coverage, the projection-slice theorem in tomography has to been introduced (Jakowatz et al. 1996). Assume that the function $g(x, y, z)$ represents the three-dimensional radar reflectivity function for a portion of the Earth's surface illuminated by the radar. The collection geometry of Fig. 11.6 shows that an azimuth angle θ and an incidence angle ψ together represent a direction from which the radar transmits a pulse and receives the return. Based on the local coordinate system (u, v, w) the projection function associated with viewing angles (θ, ψ) along the line u, slant range, is the integration of $g[x(u, v, w), y(u, v, w), z(u, v, w)]$ in the (v, w) plane, e.g. as

$$p_{\theta,\psi}(u) = \iint g\big[x(u,v,w), y(u,v,w), z(u,v,w), \big]\, dv\, dw. \tag{17}$$

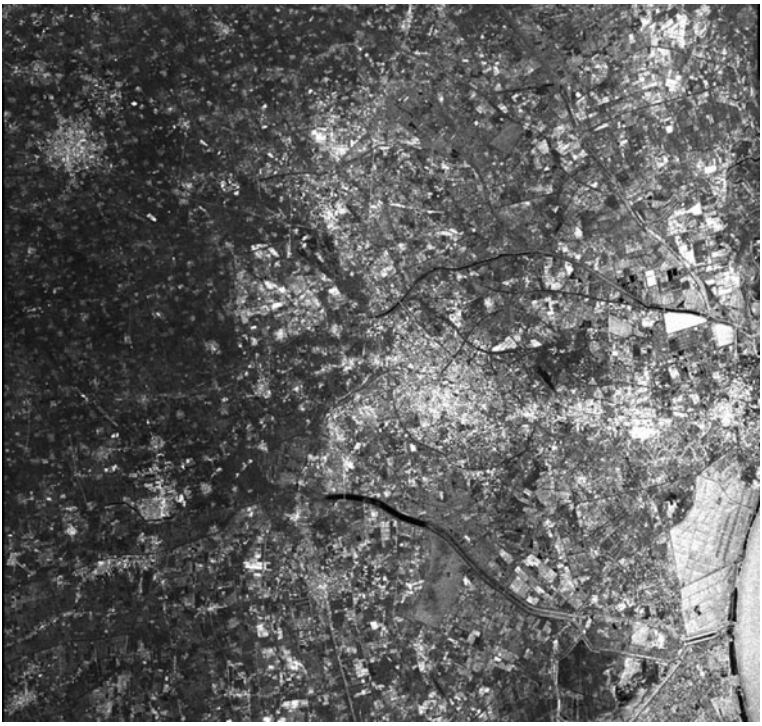


Fig. 11.4 ENVISAT ASAR strip-mode image of the city of Tianjin in China

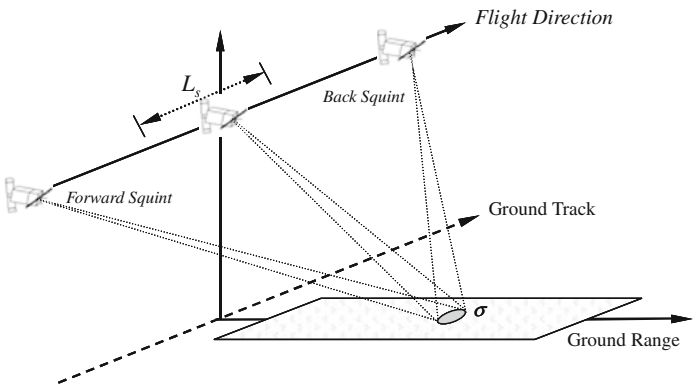


Fig. 11.5 Spotlight-mode SAR geometry

The waveform transmitted by the radar is the linear FM pulse with the carrier frequency ω_0, the chirp duration τ_p and the FM chirp rate k, as

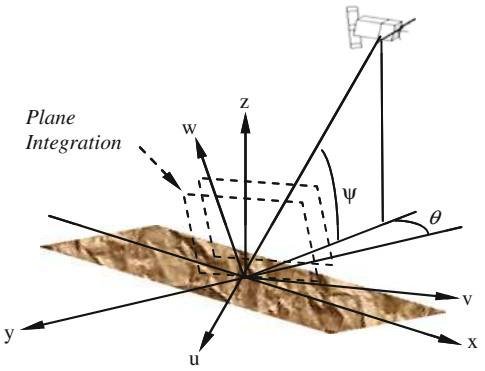


Fig. 11.6 Spotlight-mode SAR collection geometry

$$s(t) = \text{rect}\left(\frac{t}{\tau_p}\right) \mathrm{e}^{j\left(\omega_0 t + \frac{1}{2}kt^2\right)}. \tag{18}$$

At the position (θ, ψ) the return signal is

$$\begin{aligned} r_{\theta,\psi}(t) &= \int_{-u_1}^{u_1} p_{\theta,\psi}(u) \times s\left(t - \tfrac{2(R+u)}{c}\right) \mathrm{d}u \\ &= \int_{-u_1}^{u_1} p_{\theta,\psi}(u) \times \mathrm{e}^{j\left[\omega_0\left(t-\frac{2(R+u)}{c}\right)+\frac{1}{2}k\left(t-\frac{2(R+u)}{c}\right)^2\right]} \mathrm{d}u \end{aligned}$$

after quadrature modulation:

$$\begin{aligned} r_{\theta,\psi}(t) &= \int_{-u_1}^{u_1} p_{\theta,\psi}(u) \times \mathrm{e}^{-j\omega_0\frac{2R}{c}} \times \mathrm{e}^{-j\frac{2}{c}\left[\omega_0+k\left(t-\frac{2R}{c}\right)\right]u} \mathrm{d}u \\ &= \mathrm{e}^{-j\omega_0\tau_0} P_{\theta,\psi}(U)\Big|_{U=\frac{2}{c}[\omega_0+k(t-\tau_0)]}, \end{aligned} \tag{19}$$

where $P_{\theta,\psi}(U)$ is the Fourier transformation of the projection function $p_{\theta,\psi}(u)$, R is the slant range from the scene centre to the radar, u_1 is the maximum slant range for any target illuminated by the beam, τ_0 is the delay term given by $2R/c$ and c is the velocity of light. The result of the slant plane collection is a set of samples lying on a polar raster imposed on an annulus in the slant plane. According to the projection-slice theorem, the return signal is a portion of the Fourier transformation of the reflection function $g(x, y, z)$. Therefore it is possible to reconstruct the reflection function $g(x, y, z)$ and to implement SAR interferometry with the phase term caused by the delay τ_0.

Concerning the slant range and azimuth resolution, consider an annulus shown in Fig. 11.7. The entire annulus is offset from the origin by an amount equal to $2\omega_0/c = 4\pi/\lambda$. The azimuth bandwidth is determined by the radius $4\pi/\lambda$ and the extended annulus $\Delta\theta$ so that

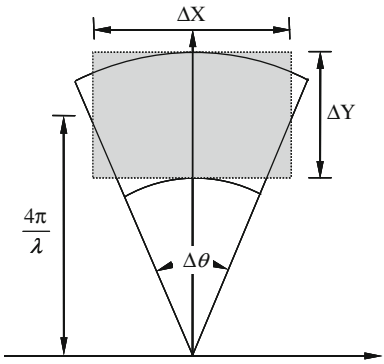


Fig. 11.7 Fourier domain polar raster samples in slant plane

$$\Delta X = \frac{4\pi}{\lambda}\Delta\theta.$$

The bandwidth in slant range is determined by impulse bandwidth B:

$$\Delta Y = \frac{2\pi B}{c/2}.$$

The resulting range and azimuth resolution are, respectively,

$$\rho_{\mathrm{x}} = \frac{2\pi}{\Delta X} = \frac{c}{2B} \quad \text{and} \quad \rho_{\mathrm{y}} = \frac{2\pi}{\Delta Y} = \frac{\lambda}{2\Delta\theta}. \tag{20}$$

It is obvious that the slant range resolution is the same as that in strip-mode SAR, but the azimuth resolution limit of $D/2$ is lifted by spotlight-mode SAR.

Figure 11.8 is a TerrarSAR-X spotlight-mode image of the Tanggu region near the city of Tianjin. The coverage is 10 km in ground range and 5 km in azimuth direction. The resolution is only 1 m. The shape of the buildings, bridges, streets, cars, ships and other constructs may be clearly recognized.

11.2.5 ScanSAR Mode

The usually used SAR imaging geometry is known as the *strip mode* with a swath width of 100 km. An important alterative mode to improve the azimuth resolution is the *spotlight mode* as introduced above. Another important and interesting mode of SAR is the *ScanSAR mode* (ESA, *ASAR Product Handbook*). In this case, the radar periodically transmits bunches of pulses also called bursts. In the time between bursts the look angle of the antenna beam is changed in order to illuminate a swath parallel to the previous one. This means that the radar is on for a period of time, then off, then on again, and so on, effectively imaging the region of interest in a

Fig. 11.8 TerraSAR-X spotlight-mode image of Tanggu acquired on 1 Jan. 2008

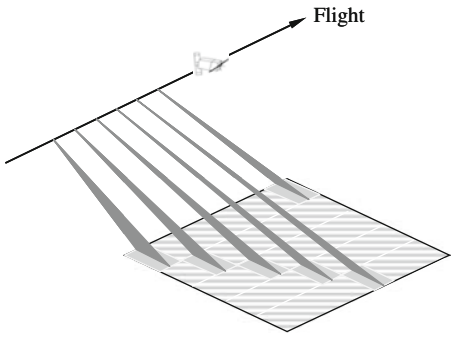


Fig. 11.9 Geometry of the typical ScanSAR mode

series of bursts, where each burst consists of a certain number of echoes. The imaging operation is split into a series of bursts of pulses, each burst providing returns from one of the subswaths, getting coverage for a strip that is 400-km wide on the ground. The ScanSAR geometry is shown in Fig. 11.9. In the case of ENVISAT, each target in the scene is observed at least three times, in three different bursts and at different times, so that the azimuth frequency of the target in the burst image is not continuous, only three portions of the true azimuth spectrum. Certainly, the azimuth resolution will be decreased because of the reductive azimuth spectrum.

Figure 11.10 is an ENVISAT ScanSAR-mode image of the area including the city of Beijing, the city of Tianjing, province Hebei and province Sandong; the coverage area is 400 km $\times$ 400 km with a resolution of 100 m acquired on 22 March 2006. The resolution of the ScanSAR image is very low, but the coverage area is very large. The images of this mode can be used when the displacement monitoring or topographic measurement for a very large area is required.

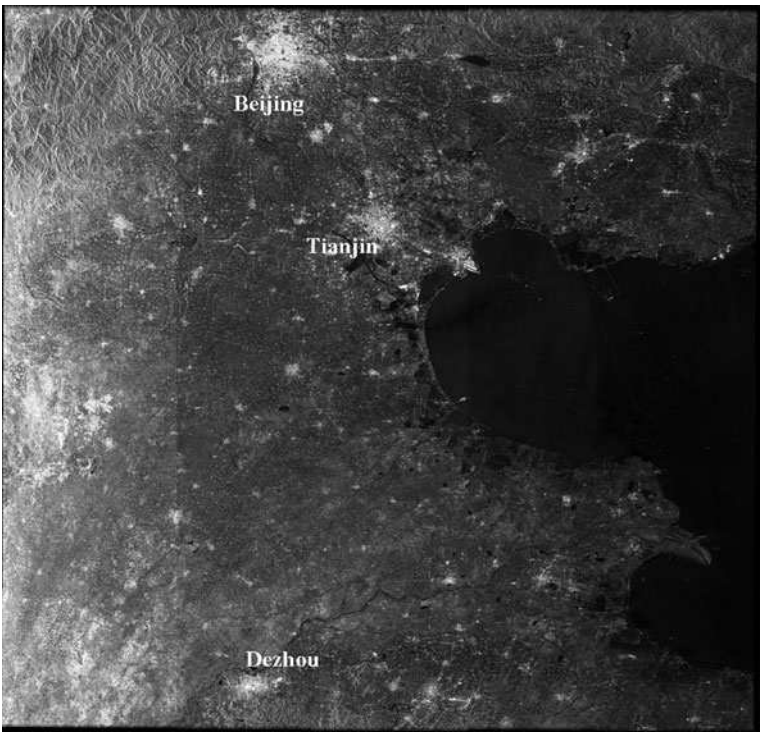


Fig. 11.10 ENVISAR SAR ScanSAR-mode image of the Beijing–Tianjin region in China

11.3 SAR Interferometry

As noted above, a synthetic aperture radar image provides the distance between the radar and the illuminated targets. As a result, an interesting possibility arises when two SAR images are produced from the same scene using very close geometry. It seems to be possible to interfere the two images in such a way and to obtain the geometric information that contains the scene topography transduced by the image data. This is the initial ideal of the SAR interferometry. In the past almost two decades, space-borne SAR interferometry was extensively developed as a powerful technique to measure the topography and deformation of the Earth's surface. In the terrain-mapping technique, the very short wavelength employed by the SAR system, some centimetres, is used as a measurement scale for the elevation derived from the SAR geometry, and therefore the accuracy afforded can be quite exceptional. Basically, there are two different configurations of the SAR interferometry system: single-pass and repeat-pass interferometric modes. In single-pass mode, two radar antennas separated by a short distance are installed on the same satellite and operated simultaneously. One transmits FM pulses to the Earth's surface and receives the return signal, while the other receives the echo pulses only. Two SAR images can

therefore be obtained at the same time. Because the return signals for the two SAR images are from the same pulses and received at the same time, they are very highly coherent and usually used for topography mapping. This mode was first proposed by L. Graham in 1974 (Graham 1974) and realized by NASA in the SRTM mission. The Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of the Earth. SRTM consisted of a specially modified radar system that flew on board the Space Shuttle Endeavour during an 11-day mission in February 2000.

In repeat-pass interferometric mode, only one radar antenna is operated, which has two jobs at the same time: transmitter and receiver. The several SAR images acquired for a same area are obtained through the repeat illuminations of the SAR at different positions and times. Although this interferometric mode can also be used for topography measurement, due to the changes of the Earth's surface with the time and the consequential temporal decorrelation, the repeat-pass mode is more suitable for deformation monitoring, for example monitoring the coseismic deformation, subsidence, landslide ice movement, etc. In the following, the principle of the SAR interferometry will first be introduced, followed then by interferometric data processing.

11.3.1 Principle of SAR Interferometry

Consider Fig. 11.11 that shows space-borne SAR interferometric geometry. Assume that P is a point target on the Earth's surface, its height relative to a reference surface is h, and acquired by the radar two times at position S_0 and S_1, respectively. Just through the two acquisitions, two SAR images are processed and presented by f_0 and f_1, master and slave image, respectively. H is the height of the satellite S_0, the distances (slant range) between the satellites S_0 and the target P are R_0 and R_1 and θ is the incidence angle of the radar. Usually the ellipsoid model surface of the Earth is used as the reference surface.

Based on the description of the SAR imaging principle in Sect. 11.2 and the INSAR geometry shown in Fig. 11.11, the SAR images f_0 and f_1 can be expressed as

$$f_i = \mathrm{e}^{-j\frac{4\pi R_i}{\lambda}} \quad i = 0, 1. \tag{21}$$

Here the reflectivity of the target is assumed to be unity. The interferometric process means that the master image is multiplied by the conjugated slave one, resulting in an interferogram g: $g = f_0 \times f_1^*$, its phase is ϕ and represented as

$$\begin{aligned} \phi &= -\frac{4\pi(R-R_1)}{\lambda} \\ &= -\frac{4\pi}{\lambda} B \sin(\theta - \alpha) \end{aligned}, \tag{22}$$

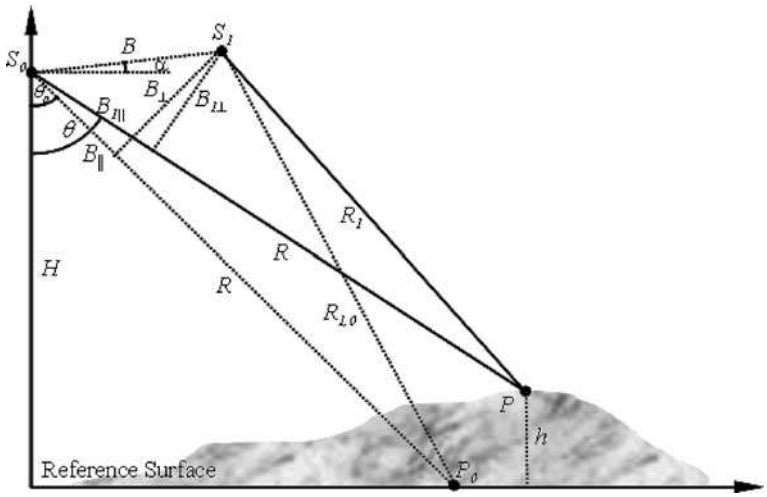


Fig. 11.11 Geometry of SAR interferometry

where B is the distance between satellite S_0 and S_1 called baseline and α is the angle of the baseline B relative to the horizontal direction. Given P_0 as a point on the reference surface, P_0 and P_1 have the equal distance R to satellite S_0, so P_0 is seen as the reference point of the point P_1 located on the reference surface. The incidence angle of the radar to point P_0 is then reduced to θ_0. Assume that the point P_0 is also illuminated by the satellite at the position S_0 and S_1, and the distance between P_0 and S_1 is $R_{1,0}$, then the interferometric phase of the point P_0 in the interferogram g is expressed as

$$\begin{aligned}\phi_0 &= -\frac{4\pi(R - R_{1,0})}{\lambda} \\ &= -\frac{4\pi}{\lambda} B \sin(\theta_0 - \alpha)\end{aligned} . \tag{23}$$

In fact, the term $B \sin(\theta - \alpha)$ in (22) is the horizontal component $B_{1||}$ of the baseline B projected on the radar–target line SP_1, and the term $B \sin(\theta_0 - \alpha)$ in (23) is the component $B_{||}$ of the baseline B projected on the radar–target line SP_0. Phase ϕ_0 is called the interferometric phase of the reference surface, also called fringe pattern, and can be accurately calculated by use of the satellite orbit vector. Subtracting ϕ_0 from ϕ, i.e. flattening the interferogram, the phase difference is

$$\begin{aligned}\Delta\phi &= -\frac{4\pi}{\lambda} B \cos(\theta_0 - \alpha) \sin(\theta - \theta_0) \\ &= -\frac{4\pi h}{\lambda R \sin\theta_0} \times B_{\perp}\end{aligned} , \tag{24}$$

or

$$h = -\frac{\lambda R \sin\theta_0}{4\pi B_{\perp}} \times \Delta\phi$$

where $B_\perp$ is the perpendicular baseline of the INSAR system. In (24) it is assumed that $R\ \sin(\theta–\theta_0) \approx PP_0 = h/\sin\theta_0$. If the interferometric phase of the reference surface is removed from an interferogram, the height of the target can be easily derived from (24). The phase of an interferogram is calculated from an argument computation; hence only principal value of phase $[-\pi, \pi)$ is generally available. Corresponding to the phase principal value, the height ambiguity $h_{2\pi}$, i.e. the height resulting in a phase change of one fringe 2π, is given by

$$h_{2\pi} = \frac{\lambda R \sin\theta_0}{2B_\perp}. \tag{25}$$

The height ambiguity $h_{2\pi}$ can be used to characterize the sensitivity of the INSAR system for terrain mapping. The larger baseline will cause the smaller height ambiguity, i.e. the higher system sensitivity. Concerning the accuracy of the topography derived from INSAR, it mainly depends on the estimation error of the interferometric phase:

$$\Delta h = -\frac{\lambda R \sin\theta_0}{4\pi B_\perp} \times \delta(\Delta\phi). \tag{26}$$

Equation (25) indicates that the larger the baseline, the better. But a large baseline will cause decorrelation of the interferometric image pair. Differentiating the interferometric phase of the reference surface in (24) with respect to incidence angle θ_0 yields

$$\begin{aligned} \delta\phi_0 &= -\frac{4\pi}{\lambda} B\cos(\theta_0 - \alpha)\delta\theta_0 \\ &= -\frac{4\pi}{\lambda} B\cos(\theta_0 - \alpha) \times \frac{\delta R}{R\tan\theta_0}, \\ &= -\frac{4\pi B_\perp}{\lambda R\tan\theta_0}\delta R \end{aligned} \tag{27}$$

where δR is the increase of the slant range R caused by the incidence angle variation $\delta\theta_0$, and approximately equals $R\tan\theta_0\ \delta\theta_0$.

The range frequency of the fringe pattern is then derived from (27):

$$\begin{aligned} f_R &= \frac{1}{2\pi} \times \frac{\delta\phi_0}{\delta R} \\ &= \frac{2B_\perp}{\lambda R\tan\theta_0} \end{aligned}. \tag{28}$$

Obviously, the local range frequency of the fringe pattern represents the frequency difference, or spectral shift, of the two SAR images in slant range, which increases with the baseline. If the spectral shift exceeds the frequency bandwidth B_R of the SAR system in slant range, the two images will be decorrelated. The range frequency bandwidth B_R of the SAR system is related to the bandwidth B of the radar

transmitted FM signal and is represented as

$$B_R = \frac{2B}{c}, \tag{29}$$

where c is the velocity of light. In the case of ESA's satellites, ERS-1/2 and ENVISAT, the radar signal bandwidth is about 15 MHz, the wavelength is 6 cm, the nominal slant range and incidence angle is 850 km and 23°, respectively, and based on these parameters the allowable baseline is about 1,100 km. Taking account of the baseline decorrelation effect and its phase sensitivity to time, the suitable baseline should be 200–600 m for the digital elevation model generation.

11.3.2 Phase Unwrapping

The phase unwrapping problem has a long history in a number of scientific research fields and applications. This problem appears in cases where a physical quantity is transduced to the phase of a complex signal. Because the phase measurement or calculation usually results from an argument computation, only the principal value of phase, or wrapped value, is generally available. For example, although phase $\phi_{(t)}$ is a linear function of time t, its value can be computed only in an interval from $-\pi$ to π, as shown in Fig. 11.12; in other words, the computed phase $\psi_{(t)}$ is modulo–2π of $\phi_{(t)}$. The process of recovering the continuous phase from its principal value is known as phase unwrapping. For a one-dimensional signal, assume that the wrapped phase array of ϕ_i is ψ_i, $-\pi < \psi_i \leq \pi$, and the goal of the phase unwrapping process is to find an integral number array k_i meeting

$$\phi_i = \psi_i + 2k_i\pi. \tag{30}$$

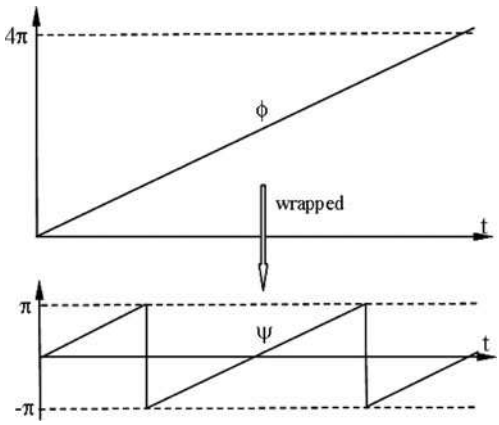


Fig. 11.12 Modulo-2π of phase ϕ

The problem is solved as follows. We begin with the first sample ψ_0, and let $\phi_0 = \psi_0$ within a constant offset, then add or subtract a multiple of 2π to the next sample, so that the absolute value of the phase difference between the second and the first sample is less than π. This procedure continues for all the available samples until the entire one-dimensional array is unwrapped. Mathematically, the phase unwrapping of the one-dimensional array is always consistently implemented, when the condition holds: all the phase differences between ϕ_{i+1} and ϕ_i are either less than π or larger than $-\pi$:

$$|\phi_{i+1} - \phi_i| < \pi \ \textit{for all } i. \tag{31}$$

Thus the phase unwrapping for the one-dimensional array can be expressed as

$$\begin{aligned} \Delta_i &= \psi_{i+1} - \psi_i \\ \phi_{i+1} &= \begin{cases} \phi_i + \Delta_i & -\pi < \Delta_i \le \pi \\ \phi_i + \Delta_i - 2\pi & \Delta_i > \pi \\ \phi_i + \Delta_i + 2\pi & \Delta_i < \pi \end{cases} \end{aligned} . \tag{32}$$

The result of (32) is always consistent; this means that the phase difference between two arbitrary points ϕ_{i+1} and ϕ_i is independent of the integral path from one point to another. This consistence is ensured because there is only one possible integral path between any two points in the unwrapped array.

The phase of an SAR interferogram is a two-dimensional matrix. In this situation the phase unwrapping process becomes significantly more complicated. Let the wrapped interferogram phase be $\psi_{i,j}$; then its unwrapped phase should be

$$\phi_{i,j} = \psi_{i,j} + 2k_{i,j}\pi k_{i,j} : \textit{integer}. \tag{33}$$

In the case of two-dimensional array, many possible unwrapping paths between any two points can be available. If the condition in (31) is still satisfied here, note that the phase distributed along any integral path becomes a one-dimensional array, so that the result of the two-dimensional phase unwrapping process is also always consistent, i.e. independent of unwrapping path. Unfortunately, in the practical situation of noise measurements, of absent data and of aliased data, it is generally impossible to obtain a consistent solution with the integral path method. Below we will introduce some methods of two-dimensional phase unwrapping for practical applications.

11.3.2.1 Least-Squares Method

Two-dimensional phase unwrapping is a most important and most difficult processing step in SAR interferometry. A least-squares solution can be obtained by minimizing the differences between the discrete partial derivatives of the wrapped phase and those of the unwrapped solution. It has been shown that this least-squares solution is equivalent to the solution of Poisson's equation on a rectangular grid with Neumann boundary conditions (Hunt 1974; Ghiglia and Romero 1989). Given

the wrapped phase $\psi_{i,j}$ on a discrete rectangular grid, our goal is to determine the unwrapped phase $\phi_{i,j}$ at the same grid locations with the aspect that the phase differences of the $\phi_{i,j}$ agree with those of the $\psi_{i,j}$ in the least-squares sense.

Define a wrapping (modulo-2π) operator W based on (33):

$$W(\phi_{ij}) = \psi_{i,j} i = 0 \ldots M-1 \text{ and } j = 0 \ldots N-1 \tag{34}$$

and

$$\begin{aligned} \Delta^x_{i,j} &= W(\psi_{i+1,j} - \psi_{i,j}) \quad i = 0...M-2, \quad j = 0...N-1 \\ \Delta^y_{i,j} &= W(\psi_{i,j+1} - \psi_{i,j}) \quad i = 0...M-1, \quad j = 0...N-2\,. \\ \Delta^x_{i,j} &= \Delta^y_{i,j} = 0 \quad \text{otherwise.} \end{aligned} \tag{35}$$

The solution $\phi_{i,j}$, which minimizes the sum

$$\sum_{i=0}^{M-2}\sum_{j=0}^{N-1}(\phi_{i+1,j} - \phi_{i,j} - \Delta^x_{i,j})^2 + \sum_{i=0}^{M-1}\sum_{j=0}^{N-2}(\phi_{i,j+1} - \phi_{i,j} - \Delta^y_{i,j})^2 \tag{36}$$

is the least-squares solution. Differentiating the sum with respect to $\phi_{i,j}$ in (36) and setting the result equal to zero yield the following system of linear equation:

$$(\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}) + (\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}) = \rho_{i,j}, \tag{37}$$

where $\rho_{i,j}$ is defined by

$$\rho_{i,j} = \Delta^x_{i,j} - \Delta^x_{i-1,j} + \Delta^y_{i,j} - \Delta^y_{i,j-1}. \tag{38}$$

It is easy to see that (38) is a discretization of Poisson's equation on an $M \times N$ grid:

$$\frac{\partial^2}{\partial x^2}\phi(x,y) + \frac{\partial^2}{\partial y^2}\phi(x,y) = \rho(x,y).$$

Equation (38) is valid in the entire rectangular grid, and the Neumann boundary condition is already included in (35).

Mathematically, a discrete Poisson's equation can be solved by means of FFTs. In order to use FFTs for solving the phase unwrapping problem (Pritt and Shipman 1994), the function $\psi_{i,j}$ has to be extended to a periodic function by performing a mirror reflection about the line$i = M$ and $j = N$ in the plane. Now applying the two-dimensional Fourier transformation on the grid defined by $0 \le i < 2M$ and $0 \le j < 2N$ to both sides of (37) yields the equation

$$\Phi_{mn} = \frac{P_{mn}}{2\cos(\pi m/M) + 2\cos(\pi n/N) - 4}, \tag{39}$$

where Φ_{mn} and P_{mn} are the two-dimensional Fourier transformations of $\phi_{i,j}$ and $\rho_{i,j}$, respectively. The expected solution $\phi_{i,j}$ is then obtained by applying the inverse Fourier transform to (39) and by restricting the result to the original $M \times N$ grid.

11.3.2.2 Branch Cuts Method

If the original scene is sampled often enough so that the true phase will not change by as much as π per sample point, the summation of the phase difference around any closed path should be zero (irrotational field). In this case, the result of phase unwrapping will be consistent, i.e. correct and independent of the integral paths. Often, in practice, this summation taking the closed multiple of 2π gives an inconsistent result around these points called residues. Assume that a closed wrapped phase path consists of n elements. The maximum gradient per element is π for all elements, except the last one having maximum value modulo-2π of $[-(n-1)]\pi$. The residue of the closed wrapped phase path will reach its maximum value as

$$|R_{\max}| = (n-1)\pi + W[-(n-1)\pi]. \tag{40}$$

Consider that an interferogram is a rectangular grid, in which the smallest closed path consists of four adjacent points. The residue is computed by (41), such that the residue may be -2π, 0 or 2π and located in the centre of the closed path shown in Fig. 11.13:

$$\begin{aligned} R = {} & W(\psi_{i,j+1} - \psi_{i,j}) + W(\psi_{i+1,j+1} - \psi_{i,j+1}) \\ & + W(\psi_{i+1,j} - \psi_{i+1,j+1}) + W(\psi_{i,j} - \psi_{i+1,j}) \end{aligned}. \tag{41}$$

Positive residues have a residual "charge" of +1 and negative residues -1 for a clockwise phase, when the residues are scaled by 2π.

Goldstein et al. (1988) introduced a phase unwrapping method to avoid the inconsistency caused by the residues. He has suggested that connection of the residues

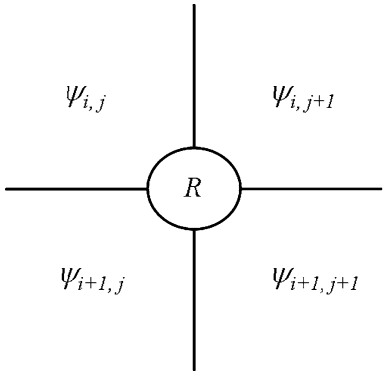


Fig. 11.13 Residue location

by "branch cuts" to create neutralized "tree" localized the phase jumps to occur across the cuts. Under ideal circumstances, these branches lie in regions that will be excluded from unwrapping. The unwrapping process begins by searching for an unvisited residue to form the start of a tree. The next step is to search other residues around each residue of the tree and then set cuts to other residues until the tree is neutralized, i.e. has the same number of positive and negative residues. At each stage the region surrounding each residue is searched up to a certain distance. Once all the residues in the current tree have been searched, the size of the search radius is increased and areas around all members of the tree are rescanned. This process continues until the tree is neutralized or the size of the search area exceeds a predetermined bound. Repeat this process to search a new unvisited residue, build the new tree and grow it. In an interferogram some trees cannot be neutralized, the reason being that the rest residues, which neutralize the tree, may lie outside the scene. In this situation set a branch cut between the tree edge and the image bound. Once all the possible trees have been constructed, the phase unwrapping is implemented by simple integration of the phase differences along any paths, which don't cross any branch cut. As a simple example, Fig. 11.14 shows an interferogram with 7×5 pixels. The phase unit is π. After the searching process (integrate the phase differences around each point), two positive (solid) and two negative (empty) residues are found and are connected by two branch cuts. The phase unwrapping can then be correctly implemented along any integral paths with the limitation that these paths don't cross any branch cut.

11.3.2.3 Minimizing Cost Flow Method

As mentioned above, branch cuts phase unwrapping method by integrating the differences of the wrapped phase along paths, avoids the regions where these phase derivatives are inconsistent. The problem of building cuts delimiting these regions is very difficult to deal with and the resulting phase unwrapping algorithm is very expensive computationally. In the least-squares method, unwrapping is achieved by minimizing the mean square differences of the derivatives between the unwrapped

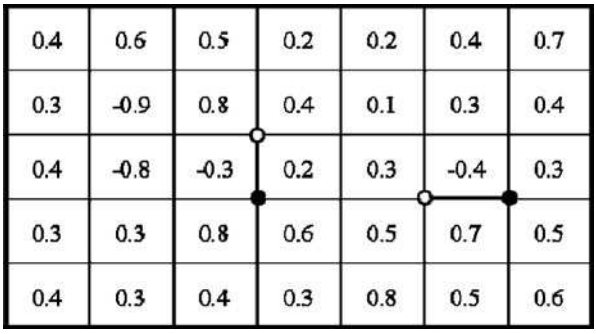


Fig. 11.14 Example of residues and branch cuts

and wrapped phases. The least-squares method is very efficient computationally, but the results are not very accurate. Costantini (1996) proposed a new unwrapping method by formulating the phase unwrapping problem as the problem of minimizing the weighted deviations between the estimated and the unknown discrete derivatives of the unwrapped phase with the constraint that the deviations must be integer multiples of 2π. This is a minimizing problem with integer variables. In this method the phase unwrapping problem is equated to the problem of finding the minimum cost flow on a network.

In this section, we combine Costantini's idea of minimizing cost flow and Goldstein's Branch Cut method to introduce a modified two-dimensional phase unwrapping method. The modified phase unwrapping method consists of three steps: constructing residue network with regular or irregular grid, building branch cuts in the grid by means of a minimizing cost flow algorithm and finally unwrapping the phase.

Constructing a residue network is the first step of the modified method. Assume that an interferogram has line number of $M + 1$ and each line has pixel number of $N + 1$. Assume also that all the residues are calculated by (22) with a scale of $1/2\pi$ and located as in Fig. 11.13. When considering all the computed residues, i.e. the entire scene, the residue network is a regular rectangular grid with M lines and N columns. The residues in the grid are also called the nodes. Each pair of two adjacent residues is connected by an arc. In order to neutralize all the residues, a common node is added to the network and connected to all the nodes lying on the two boundary lines and two boundary columns. All the nodes and arcs compose the residue network. Positive residue charge (+1) means a supply node and negative residue charge (-1) means a demand node. Based on the definition above, in the constructed network the number of nodes is $M \times N + 1$ and coded from $Node_0$ to $Node_{M\times N}$, and the number of arcs is $2MN + M + N - 4$ and coded from Arc_1 to $Arc_{2MN+M+N-4}$. Each node has four associated arcs, but corner node has only three. Each arc has two associated nodes: start node and end node. Moreover, it should be noticed that the supply or demand of the common node is dependent on the difference of the total supplies and total demands. The regular residue network is shown in Fig. 11.15, in which all the boundary nodes are connected to the added node $Node_0$ that is not shown in the figure.

The second step is building the branch cuts by means of network programming, i.e. minimum cost flow approach. As each supply has to be transported from a supply node to a demand node along a path consisting of several arcs, we can let C_i and S_i be the unit cost and shipment or flow on the Arc_i, respectively, so that the total cost is represented as

$$\sum_{i=0}^{I} C_i S_i, \tag{42}$$

where $I = 2MN+M+N-4$. Minimizing (42) under the constraint that each supply must be transported to a demand node, and all the demands must be satisfied, we

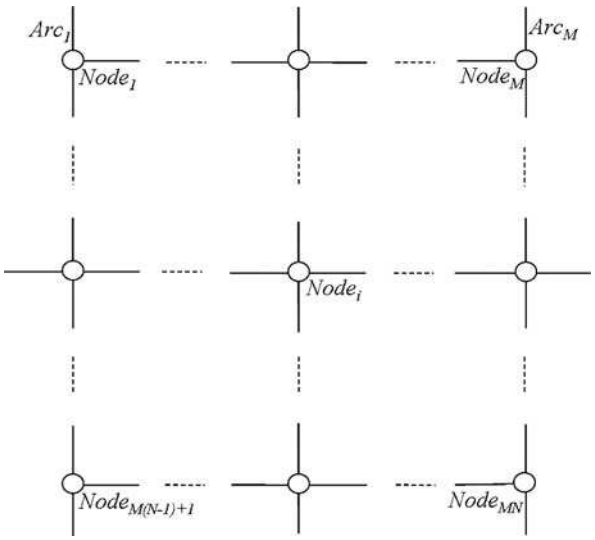


Fig. 11.15 Regular residue network

can obtain the shipment distribution of the network with the minimum cost. The collection of all the arcs with non-zero shipments is therefore a realization of the expected branch cut trees. At the end of the second step all the positive and negative residues are properly connected.

Finally, in the third step, we can unwrap the phase from any point to another by integrating the wrapped gradients along any path connecting the two points and not crossing any cut.

11.3.3 Image Registration

The interferogram is the interferometric result from a complex SAR images acquired by radar for a same area but at different radar positions, such that the image coordinates of a same target in the two images are different. This means that there is an offset between the two images. After analysing the SAR geometry, we know that the offset, i.e. the image coordinate difference of same target, depends on the SAR geometry, satellite orbit, radar timing parameter and the topography of the target. In order to generate a correct interferogram the slave image must be registered to the master image, such that each target in the two images has identical coordinates (exactly same range and azimuth position). The aim of the coregistration is to find a transformational model between master and slave image. Usually, the four-coefficient polynomials used to model the range and azimuth offsets are expressed as

$$\begin{aligned}\Delta R_{i,j} &= a_0 + a_1 \times i + a_2 \times j + a_3 \times i \times j \\ \Delta A_{i,j} &= b_0 + b_1 \times i + b_2 \times j + b_3 \times i \times j\end{aligned}, \tag{43}$$

where $\Delta R_{i,j}$ and $\Delta A_{i,j}$are the range and azimuth offsets of the target located at (i, j) in the master image, respectively, i is azimuth line number and j is range pixel number.

Usually, conventional techniques to estimate the offset vector $(\Delta R_{i,j}, \Delta A_{i,j})$ are based on the cross-correlation of the amplitude images with a moving window. Oversampling of the slave image is required for a coregistration accuracy of better than 0.2 pixels. Due to temporal decorrelation caused by surface change, the cross-correlation degree is very low sometimes or in some local regions. Therefore, over the entire image, for a large number of windows, the offsets are estimated in least-square sense and only those cross-correlation results with higher degree are used for solving (43).

After computing the range and azimuth offsets, the slave image is then resampled with the offset estimation; this process is also called image fitting. Because of the variation of the offset with the pixel position, the resampling accuracy is also necessarily to be guaranteed.

11.3.4 Coherence of SAR Images

A condition guaranteeing the implementation of SAR interferometry and the accuracy of its result is that the master and slave images must be highly coherent. The coherence is mainly regarded as an estimation of the phase stability of the illuminated targets during the two acquisitions. Because the two images are acquired by radar at different time spanning a few days or months, even several years, the surface changes have certainly appeared during the time span resulting in the low image coherence and decrease the interferometric measurement accuracy.

The normalized coherence γ is defined as the complex correlation between two co-registered complex images:

$$\gamma = \frac{\langle f_1 \times f_2^* \rangle}{\sqrt{\langle f_1 \times f_1^* \rangle \langle f_2 \times f_2^* \rangle}}, \tag{44}$$

where the brackets <> indicate the ensemble average and * denotes the conjugate multiplication. Equation (44) is only a definition of coherence. In practice, the coherence is estimated by averaging the complex value over a window with a finite size and is shown in the following equation:

$$\hat{\gamma} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} f_1 \times f_2^*}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} f_1 \times f_1^*} \times \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} f_2 \times f_2^*}}. \tag{45}$$

Usually, the absolute value of $\hat{\gamma}$ is used and gives coherence value in the range [0, 1].

11.4 Differential SAR Interferometry

It is well known that topography is the interferometric result of an SAR image pair, i.e. derived from an interferogram. This SAR image pair is acquired over the same region and at the same time (single-pass INSAR). If the two SAR images are acquired at different times (repeat-pass INSAR) and in order to obtain the true topography, the radar-illuminated surface has to be absolutely stable, no changes appearing during the two acquisitions. However, the reality is often not the case. Some human activities (construction project) or natural factors (geologic hazard) will cause changes to the Earth's surface. Certainly, if an interferogram simultaneously contains information on both topography and changes (displacement or movement), and in some way, the topography may be subtracted from the interferogram, for example using another interferogram containing topography only, or using an existing digital elevation model, the changes can be retrieved. This process can also be seen as interferometry of two interferograms and is called differential SAR interferometry (D-INSAR).

11.4.1 Principle of D-INSAR

Consider Fig. 11.11 again. Sometimes the point P is naturally or artificially moved to P' and presents on the second slave image f_2 shown in Fig. 11.16. At this time, the radar is located at S_2, and the distance between P' and the radar becomes $R_2+\rho$; here R_2 is the distance between S_2 and point P, and ρ is the increment of the slant range. In the second interferogram $g_2 = f_0 \times f_2^*$ consisting of the master and the second slave image, the phase of the point P' is

$$\Phi_2 = -\frac{4\pi(R_0 - R_2 - \rho)}{\lambda} = -\frac{4\pi(R_0 - R_2)}{\lambda} + \frac{4\pi}{\lambda}\rho. \tag{46}$$

Obviously, the contribution of the movement ρ in the interferogram is independent of the baseline. As with the first interferogram, the second interferogram also includes information of the phase of the reference surface. Subtracting the phase of the reference surface, the topography and the movement components remain:

$$\Delta\Phi_2 = -\frac{4\pi h}{\lambda R \sin(\theta_0)} B_{2,0\perp} + \frac{4\pi}{\lambda}\rho. \tag{47}$$

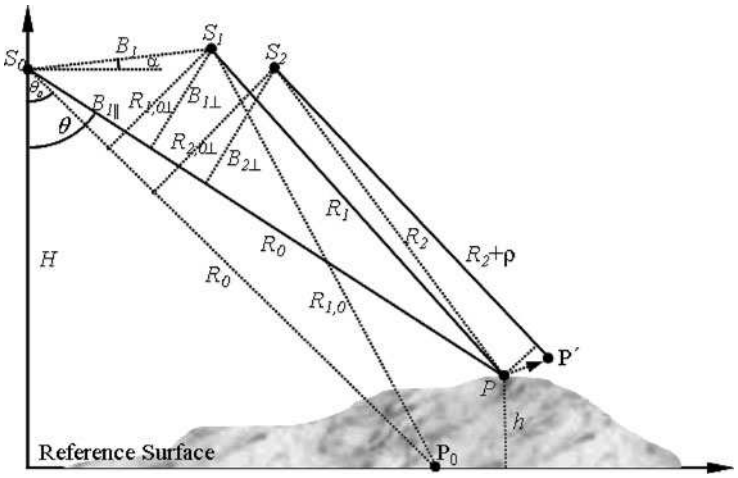


Fig. 11.16 Geometry of differential SAR interferometry

Consider the topographic components in the two interferograms (24) and (47); their difference only lies in their different scales. If the first interferogram (24) is multiplied by a factor $B_{2,0\perp}/B_{1,0\perp}$ and subtracted from the second interferogram (47), the movement component can be obtained:

$$\frac{4\pi}{\lambda}\rho = \Delta\Phi_2 - \frac{B_{2,0\perp}}{B_{1,0\perp}}\Delta\Phi_1, \tag{48}$$

which is the basic principle of differential SAR interferometry for detecting a small movement or displacement. Here the moving ρ is the projection of the true movement on the radar–object line only. To find the true movement vector some preexistent knowledge or other additional measurement is needed. The phase component of the topography in (47) can also be extracted from other existing DEM, for example from SRTM and precise orbit data. For each pixel of an interferogram the corresponding satellite orbit time and range time are known, and the satellite position and distance between the satellite and the pixel can be calculated by use of precise orbit data. Therefore, the phase component of the topography may be synthesized and then removed from the interferogram.

11.4.2 Persistent Scatterer SAR Interferometry

As described above, the phase of a pixel in an interferogram indicates the summed result of topography and displacement of an observed target, associated atmospheric delay and other changes at the time of the acquisitions. Their individual phase components are superimposed in the interferogram. In some cases, the phase components of other changes can be practically neglected or independently estimated,

and the interferometry is therefore successfully applied. However, in most cases, due to temporal and geometrical decorrelation, the phase unwrapping process is very difficult, usually even impossible. The permanent scatterer (PS) technique has been developed by Ferretti et al. (2001) to overcome these problems. PS technique utilizes all archived but suitable data of a certain area, stacks differential interferograms co-registered to a common master one. Instead of analysing the phase in the spatial domain in the entire scene, the phase of isolated coherent points is analysed as a function of time and space. The PS technique is protected by a patent of Ferreti et al. and the term "Permanent Scatterer Technique" is trademarked. Therefore many scientists use another term – "Persistent Scatterer Technique" – with the same idea but with some small differences (Werner et al. 2003).

11.4.2.1 Selection of Master Image

All the available images on the same track and frame must be co-registered to a master image. The master image is selected for decreasing the dispersion of the perpendicular baselines of all the possible interferometric combinations, and the acquisition time of the master image lies in or near the middle of the entire acquisition time range. In this chapter only those images from same sensor will be considered, so that the difference of Doppler centroid frequency is very small and therefore will not be considered. When the master image has been chosen, all the interferograms between the master image and slave images will be generated. Before the interferogram generation all the images including master and slave images are oversampled by a factor of two in range and azimuth direction in order to avoid aliasing of the complex interferometric signal.

11.4.2.2 Generation of Differential Interferograms

A reference digital elevation model (DEM) and precise orbit data are used to obtain the simulated interferometric phase induced by topography. The differential interferograms are then generated by subtracting the simulated interferometric phase of topography from the interferograms. In fact, the interferometric phase of topography can be computed from a subset of the available images, preferably with large perpendicular and small temporal baselines, if these images are highly coherent everywhere.

11.4.2.3 Modelling of Differential Interferometric Phase

The differential interferometric phase for each point in the kth differential interferogram, if unwrapped, can be modelled (Werner et al. 2003; Colesanti et al. 2003) as

$$\phi^k = \phi^k_{\text{defo}} + \phi^k_{\text{topo}} + \phi^k_{\text{atmo}} + \phi^k_{\text{noise}}, \tag{49}$$

where ϕ^k_{defo} is the phase caused by point deformation, ϕ^k_{topo} is the phase due to inaccuracy of the reference DHM, ϕ^k_{atmo} is the phase induced by atmospheric delay and the last term ϕ^k_{noise} is the phase due to decorrelation and other noise. The topographic phase is a linear function of the perpendicular baseline and can be expressed as

$$\phi^k_{\text{topo}} = \frac{4\pi B^k_{\perp}}{\lambda R \sin\theta_0} \times \Delta h, \tag{50}$$

where Δh is the height of the point relative to the reference surface, referred to as DEM error. The phase ϕ^k_{atmo} is a random variable in the temporal domain and independent of the INSAR system, but can be neglected by considering phase difference between nearby points. The noise term is a high-pass signal and contains all other phase contributions. The displacement component for each point can be considered as a linear function of the acquisition time. If the acquisition time difference of the kth differential interferogram is T^k and the average displacement rate of the point is ρ, the displacement phase can be written as

$$\phi^k_{\text{defo}} = \frac{4\pi}{\lambda} \times T^k \times V. \tag{51}$$

The estimation of the average displacement rate V and DEM error Δh is not accomplished for the entire scene, but only for those coherent points with point target like scattering characteristics, e.g. the persistent scatterers.

11.4.2.4 Preliminary Estimation of Persistent Scatterer Candidates (PSCs)

The permanent or persistent scatterers in the differential interferograms should be those points which have very small phase dispersion with time. The process identifying the PS points by analysing their phase behaviour only is not possible because the phase still contains unknown signal contributions. Ferretti et al. (2001) have shown that the estimation of the phase stability on the amplitude dispersion holds very well for standard phase deviation $\sigma_\phi < 0.25$ radius resulting in a numerical simulation using 5,000 series with 33 SAR images. An amplitude dispersion index D_{a} is defined as a ratio of the amplitude standard deviation σ_a over the amplitude mean $\overline{a}$ for certain points and equals the phase standard deviation σ_ϕ:

$$\sigma_\phi = \frac{\sigma_a}{\overline{a}} = D_a. \tag{52}$$

This relation means that a point having similar, relatively large, amplitude during all acquisitions is expected to have a small phase dispersion. This enables detection of isolated points, which is not possible if the detection is based on a spatially estimated coherence value. The points with a smaller amplitude dispersion index are considered to have smaller phase dispersion. Typically, the threshold of D_{a} is set to be about 0.25. Thus, a point is selected as a candidate if its amplitude dispersion

index is below the threshold. In order to ensure that the amplitude dispersion index is correctly estimated, more than 30 images are needed; this condition usually can't be satisfied in most areas in the world.

Werner has suggested a method to generate persistent scatterer candidates with "point target characteristics" in her INSAR software "gamma". This is implemented based on a single SLC image and based on the spectral diversity. The idea of Werner is that the energy of a point target remains more or less the same when processing different looks with fractional azimuth and range bandwidth. Based on this low spectral diversity, respectivity based on correlation values above the indicated threshold potential point targets can be identified. The physical basis for the methodology used is that point targets do not exhibit the speckle observed for extended targets. For a point target almost the same backscattering intensity is found when processing different looks with fractional azimuth and range bandwidth. The mean to standard deviation ratio (MSR) between the spectral looks is used as the (inverse) measure of the spectral variability. Lower spectral variability corresponds to higher MSR values and 1.2 might be a possible threshold to use. For a series of co-registered SLC all the selected point candidates have to be wise-merged.

Before detecting PS candidates, all the images have to be radiometrically calibrated in order to ensure that the calibrated images are comparable and allow for correct estimation of σ_a and $\bar{a}$. The calibration is usually based on annotated radar system parameters given in the leader file. In practice, the histograms of the intensity images can also be used for the calibration process. The variation of the modes of the histograms must be less than an expected value, for example 1 dB. Otherwise the histograms should be shifted by multiplication of the intensity with a proper constant to ensure that the difference between the modes of the histograms is small enough. The histogram should be calculated in a high coherence area to avoid random changes of the illuminated scatterers.

11.4.2.5 Estimation of Linear Deformation

As the phase of the differential interferogram is wrapped, the estimation of the displacement rate and the DEM error is a non-linear inversion problem. Assume that there are N differential interferograms and each PS candidate has a phase series with N elements. Starting from a previously known true persistent scatterer n, the following complex ensemble coherence in time is used as a measure of the phase stability of a nearby PS candidate m and hence an indicator of whether the candidate is a true PS relative to the true PS n:

$$\gamma_{m,n} = \frac{1}{K}\sum_{k=1}^{K} \exp\left\{ j\left[\phi_{m,n}^{k} - \frac{4\pi}{\lambda}\left(T^{k} V_{m,n} - \frac{B_{\perp}}{R \sin\theta_0}\Delta h_{m,n} \right) \right] \right\}. \qquad (53)$$

Here n indicates a true PS near the PS candidate m, $\phi_{m,n}^{k}$ is their phase difference, $V_{m,n}$ is their displacement rate difference and $\Delta h_{m,n}$ the difference of the DEM error. All the terms related to atmospheric delay and noise approximately are zero by assuming that the nearby points (distance is not larger than 1 km) have the same atmospheric delay, and noise is a white random noise in time, too. The amplitude

of the complex coherence is less than or equal to 1, and therefore if a candidate has the coherence value over a threshold, say 0.8, by maximizing the absolute of $\gamma_{m,n}$ with the arguments $V_{m,n}$ and $\Delta h_{m,n}$, this point is determined as a true PS with the estimated $V_{m,n}$ and $\Delta h_{m,n}$. All the currently existing true PS are then used for detecting other true PS and their displacement rate and DEM error difference.

11.4.3 Example: Coseismic Deformation Measurement of Bam Earthquake

On 26 December 2003 an earthquake (Mw = 6.5) shook a large area of the Kerman province in Iran. The epicentre of the devastating earthquake was located near the city of Bam. Almost all the buildings of Bam were damaged and over 30,000 people died. This example described the application of differential synthetic aperture radar interferometry (D-INSAR) and ENVISAT ASAR data to map the surface deformation caused by the Bam earthquake including the interferometric data processing and results in detail (Xia 2005; Wang et al. 2004). Based on the difference of the coherence images before and after the event and edge search of the deformation field, a new fault ruptured on the surface was detected and used as a data source for parameter extraction of a theoretical seismic modelling. The simulated deformation field from the model perfectly coincides with the result derived from the SAR interferometric measurement.

11.4.3.1 Radar Data

The European Space Agency (ESA) supplied an ENVISAT ASAR data set on Bam area including four descending pass single look complex images: orbits 6687, 9192, 9693 and 10194, acquired on 11 June 2003, 3 Dec. 2003, 7 Jan. 2004 and 11 Feb. 2004, respectively, and three ascending pass SLC images, acquired on 16 Nov. 2003, 25 Jan. 2004 and 29 Feb. 2004, respectively. In the descending pass data set, the first two passes are acquired before the event, while the last two passes are acquired after the event; therefore each pair can be used to generate a digital elevation model. Any combination of an image acquired before the earthquake and another one acquired after the earthquake can be used for measuring the coseismic displacement information caused by the Bam earthquake on the satellite-object line. Because the Doppler frequency centre shift of the data acquired on 25 Jan. 2004 is too large, this image is not used in this chapter. All the data are sorted by orbit direction and co-registered before data processing.

11.4.3.2 Baseline Estimation

The baseline estimation is very important in the INSAR data processing. It may be known by the front analysis that the phase of a pixel in an interferogram equals the sum of the reference points' interferometric phase on the reference plane and the

phase produced from the point's height, if we don't take the movement into account. The two addends are even scaled by the horizontal and the vertical components of the baseline. It should be noted that the point on the surface and its reference point on the reference plane are not the same, but they appear at the same position in the master image. Therefore, the baseline estimation needs to calculate the horizontal and vertical baseline for each pixel and suppose that all the pixels lie on the reference plane. In the master image, each azimuth line corresponds to a determinate position of the satellite. The azimuth line is defined in the master image. In the co-registered slave image, the pixels of an azimuth line may come from several original lines. In other words, an azimuth line in the co-registered slave image should correspond to different positions of the radar in flying. Because of the difference of the depression angle and distance of the radar for each point on an azimuth line and the non-parallelity of the satellite orbits, every pixel on a same azimuth line has its own horizontal and vertical baseline and the variation is in addition not linear. Some of the literature said that the baseline for an azimuth line is the same and perpendicular to the orbit but this is not correct. During the flying of the satellite for a standard ENVISAT ASAR product, usually 16 s, the variation of the baseline may reach several metres, even more than 10 m. The ENVISAT ASAR product also provides the corresponding satellite orbit data and other auxiliary data but no precise orbit data such as ERS-1/ERS-2, its former generation, which provided ERS precise orbit data with an accuracy of some 10 cm (Reigber et al. 1996). For the interferogram generation the baseline should be corrected and refined.

In fact, the estimation of the baseline needs to know the coordinates of all the pixels on the reference surface and the corresponding satellite's space locations. Based on the satellite orbit data and the associated acquisition time, one can first determine the space coordinates of the satellite for each azimuth line in the master image, then the coordinates of each pixel on the azimuth line. Usually the surface of an ellipsoid model of the Earth is selected as the reference surface (for example WG84); this means that all the reference points have to be located on the ellipsoid model surface. The distance between the reference point and the satellite depends on the width of the radar's receiving time window and is perpendicular to the satellite orbit. Each pixel is located at the position when its Doppler frequency is zero. Under the conditions mentioned above, the coordinates (x_p, y_p, z_p) for each pixel have to satisfy the following equation system (Curlander 1982):

$$\begin{cases} \dfrac{\left(x_p^2 + y_p^2\right)}{R_e^2} + \dfrac{z_p^2}{R_p^2} = 1 \\ \dfrac{2}{\lambda R_s}(\vec{V}_s - \vec{V}_p) \times (\vec{S} - \vec{P}) = 0 \\ \sqrt{(\vec{S} - \vec{P}) \times (\vec{S} - \vec{P})} = R_s \end{cases} \tag{54}$$

In this equation system, R_s is the distance between satellite and reference point, R_e and R_p are the major and minor axis of the Earth ellipsoid, V_s is the satellite's velocity vector, V_p is the velocity vector caused by the Earth's rotation, S is the coordinate

Table 11.1 Interferometric pairs of ENVISAT data and corresponding baseline

Orbit number	Acquisition date	Orbit direction	Baseline $B_\perp$ (m)
9192/6687	3 Dec. 2003/11 June 2003	Descending	484.1 to 468.6
10194/9693	11 Feb 2004/7 Jan. 2004	Descending	–540.6 to –508.3
9192/10194	3 Dec. 2003/11 Feb. 2004	Descending	3.1 to 3.4
10459/8956	29 Feb. 2004/16 Nov. 2003	Ascending	–13.6 to –20.7

vector of the satellite and P is the coordinate vector of a pixel (x_p, y_p, z_p) on the reference surface. After the coordinate vector (x_p, y_p, z_p) is solved, one can draw a perpendicular line from the reference point to the orbit of the slave satellite and get the coordinate vector of the slave satellite. If all the coordinate vectors of the satellite and reference points are known, the baseline may be easily and directly computed and then applied in the flat removing and differential interferometric process; see (44) and (48).

Table 11.1 shows all the interferometric pairs with their orbit number, acquisition date and baseline estimation for the near range and far range.

11.4.3.3 Interferogram and Elevation Model

Figure 11.17a shows the radar intensity image of the city of Bam and its surrounding (radar coordinate) with an area of 42 km× 42 km. The city of Bam is located at the left upper side of the image centre. Figure 11.17b, c shows two interferograms after flattening and are derived from the data pair 9192/6687 (before event) and 10194/9693 (after event). Each fringe indicates a phase variation of 2π or a height variation of about 17 and 19 m, respectively. Figure 11.18a, b shows the integrals of Fig. 11.17b, c, namely the results after phase unwrapping. They represent the accumulation of the phase variation, or say, a scale of the true topography. Figure 11.18c is the differential result of Fig. 11.18a, b. It is not necessary to transfer the phase into the height in the differential process; see (48). The homogeneous result of the differential process proves that the topography in this area doesn't change before the earthquake or after the earthquake. Both images, Fig. 11.18a, b, may be used as an elevation model. In addition, the homogeneous result of the differential process (Fig. 11.18c) proves that the integral of Fig. 11.17b, c has no error and the baseline estimation is also correct. In all the three images in Fig. 11.18, the phase is represented with colour, but the intensity and the saturation are modulated with the radar intensity and the coherence value for highlighting the information of the relative location.

11.4.3.4 Differential Interferometry and Surface Deformation

If an interferogram is calculated from two scenes of radar data acquired, respectively, before and after the earthquake, this interferogram contains the topography information and moreover includes the deformation information caused by the

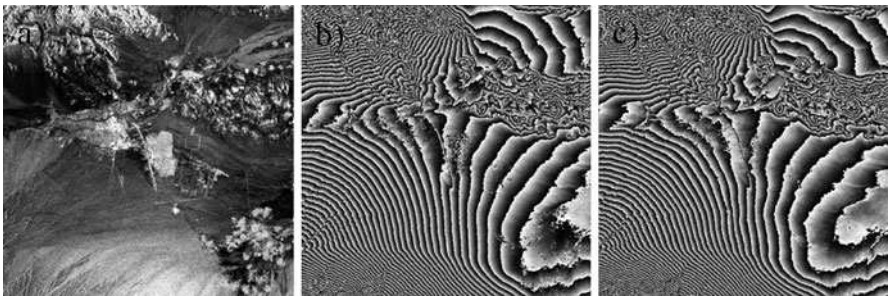


Fig. 11.17 (**a**) Radar intensity map. (**b, c**) Interferograms after flattening

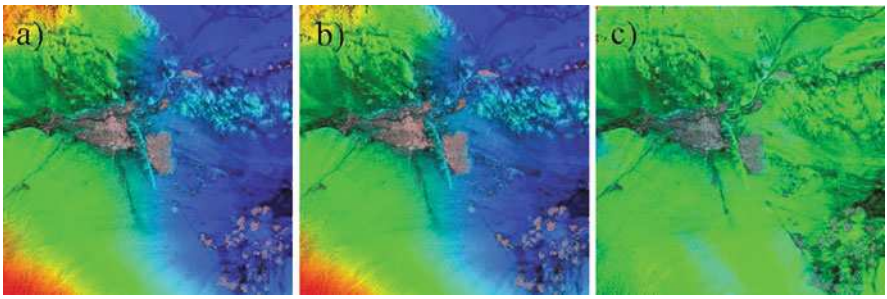


Fig. 11.18 (**a, b**) Unwrapped phase of the interferograms. (**c**) Phase difference of (**a**) and (**b**)

earthquake. Using the orbit 9192 and 10194 as an interferometric pair, the corresponding acquisition date is 3 Dec. 2003 and 11 Feb. 2004; subtracting the topography represented in Fig. 11.18a from this interferogram, the phase due to the deformation remains in the result of the differential interferometry and is shown in Fig. 11.19a after geocoding. In Fig. 11.19a the topography is all subtracted, and only the phase variation from the deformation remains. A colour period still indicates a phase variation of 2π, but in this case, the phase variation means the deformation, which equals half wavelength, 2.8 cm: see (47). The deformation appears mainly in the city of Bam and in the surrounding of about 25 km apart from the city. As mentioned earlier, the deformation measured from the D-INSAR is the projection of the true deformation vector on the line of satellite (LOS) only. With respect to the descending orbit of ENVISAT and the Bam region the direction cosine vector is (–0.066, 0.384, 0.920). In the deformation field only two plum blossom petals may be seen, half of four, which are located in the right side of the image, one in the north and one in the south. The south petal lifts 30 cm, while the north one goes down 18 cm. Figure 11.19b is the deformation field derived from the ascending orbit data pair 10459/8956. The corresponding direction cosine vector is (–0.057, –0.332, 0.492). In Fig. 11.19b three plum blossom petals may be seen, but the amplitude is relatively small. The northwest petal lifts 12 cm, the northeast one goes down 6 cm, while the southeast one lifts 18 cm. Because the position of the southwest petal

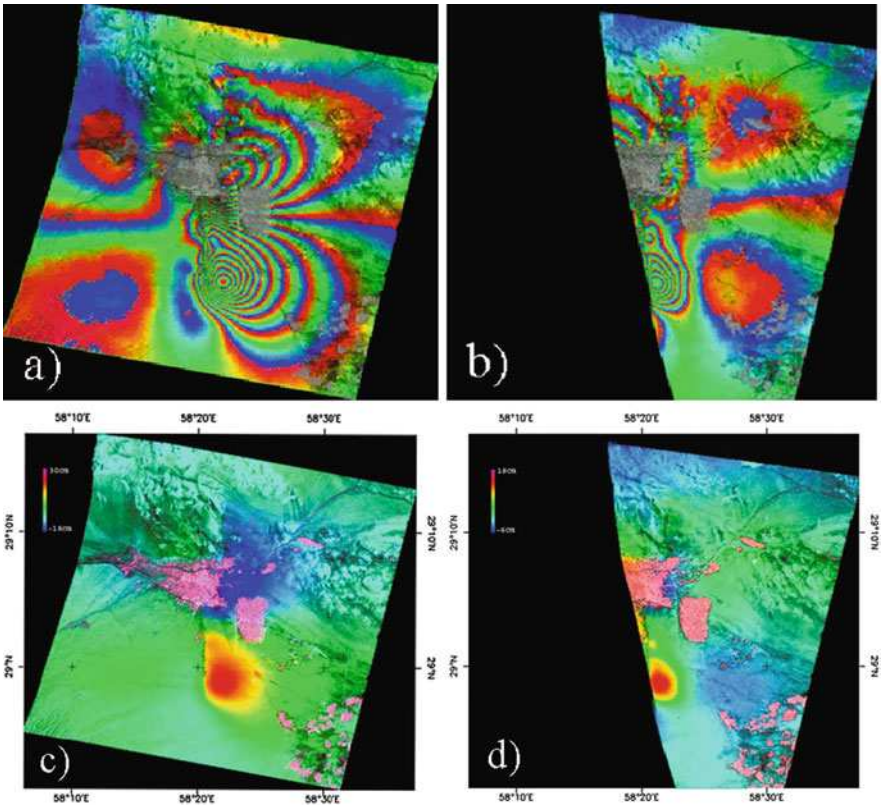


Fig. 11.19 Coseismic deformation: (**a, b**) before phase unwrapping; (**c, d**) after phase unwrapping

exceeds this ascending image, only a small arc of it can be seen, but shows that it is down. Analysing the distribution of the uplifting and subsiding of all the petals, one may conclude that the fault should be south-north trend. The distribution of the plum blossom petals indicates that this event is a right-lateral strike-slip earthquake. The fault seems to consist of several segments. The main portion of the deformation arises from the relative horizontal south-north slip, which is even perpendicular to the LOS in the space. This makes the contribution of the horizontal displacement in Fig. 11.19a, b negligible. Figure 11.19c, d shows the integrals of the images at Fig. 11.19a, b, namely the deformation after phase unwrapping and marked with the geographical coordinates (geocoded). The deformation appears mainly near the city of Bam and has reduced to 0 besides 25 km apart from the epicentre.

11.4.3.5 Determination of the Location and Shape of the Ruptured Fault

The Bam earthquake arose because of a fault break. If there had also arisen a fissure (a type of damage) on the ground, the relative displacement quantity of the points

nearest the fault should be the greatest. Thus, the location and shape of the fault may be determined by means of detecting the fissure. The first method is to compare the three coherence images: acquired before the earthquake, after the earthquake and before/after the earthquake. The fissure may appear only in the coherence image derived from the data pair acquired before and after the earthquake. Because of limited space, here only the third coherence image is given and used for the colour modulation in Figs. 11.18 and 11.19. In these images the dark line in the north and south of the city of Bam indicates the new fissure. In the coherence image before or after the earthquake this dark line does not exist. In the city of Bam, the fissure cannot be detected since the coherence value here is too small. It may be noted that the new fissure must cross the city. The second method is the use of the Edge Detection Filter. Figure 11.20 shows the output of the filtering process. The shape and the location of the fault are revealed very clearly. Evidently it is roughly from 3 segments, the southern segment, the northern segment and the middle one. The middle segment in the city of Bam is out of order. The locations of the end points of the northern section are 58°21′57.68″E/29°11′36.00″N and 58°22′10.29″E/29°7′36.78″N, about 6 km; the locations of the end points of the southern segment are 58°20′19.91″E/29°4′57.30″N and 58°20′45.14″E/28.58′27.46″°N, about 13 km. The middle segment disappeared below the city area of Bam, where a lack of coherence prevents its tracking. This segment may be imagined as a join line from the southern end of the northern segment to the northern end of the southern segment, about 5 km.

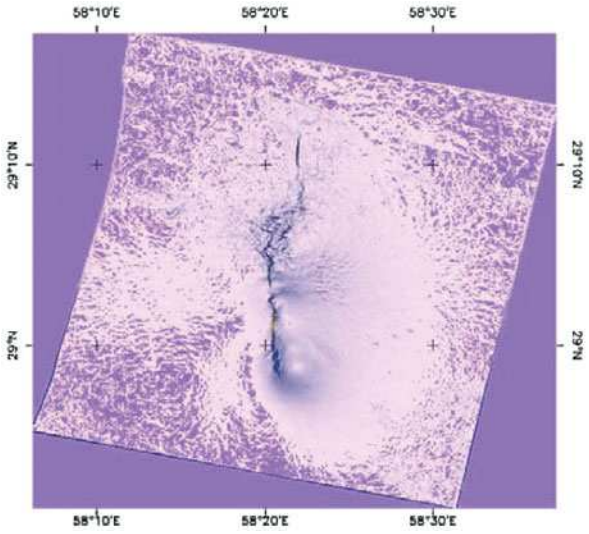


Fig. 11.20 Detection of ruptured fault

11.4.3.6 Estimation of the Theoretical Model for the Earthquake Source

The deformation field, the shape and the location of the fault measured from the differential SAR interferometry may help geophysicists to deepen their research and simulation for the theoretical modelling of the earthquake source. Dr. Wang Rongjiang, scientist at the Geo-Research Center, Potsdam, estimated the earthquake source parameters using the elastic half-space theoretical model provided by Prof. Okada (Japan) and the deformation field derived from this chapter as well as the least-squares method. Figure 11.21a, b shows the modelling results for the surface deformation of the Bam earthquake from the descending and ascending data, respectively. Comparing Fig. 11.21a with Fig. 11.19a and Fig. 11.21b with Fig. 11.19b shows that they are very consistent. Variance is around 2 cm; greater discrepancy is near the fault only and is only a few centimetres. It is possible that the theoretical model and parameters cannot completely describe the tiny structure of the fault. Referring to the geological map of the Bam region, it may be seen that the northern section of the fault coincides with an old fault (Bam fault), although it does not extend like the old fault to the southeast, but turns to the southwest, crosses the city of Bam, and then extends south, forming a brand new southern fault. This brand new fault is nearly parallel to the old Bam fault and approximately 5 km away. The simulated result of Dr. Wang shows that the length and the trend of the northern, middle and southern fault plane are 7 km/0°, 5 km/35° and 14 km/357°, respectively.

The northern section of the fault inclines 55° to the west, the slip is very little, about 20–70 cm, while the middle and southern sections incline to the east near 80°, the southern section slips 1–2 m, maximally reaching 2.75 m. The middle one is a transitional zone. Evidently, more than 80% of the seismic moment was released from the southern section and the epicentre should be in the southern section. This is also one of the major reasons for the crushing damage in the city of Bam.

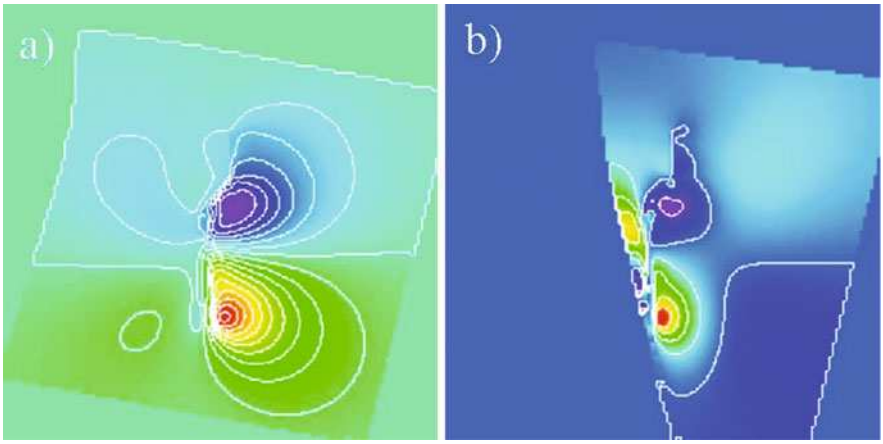


Fig. 11.21 (**a, b**) The modelling results for the surface deformation of the Bam earthquake from the descending and ascending data

11.4.4 Example: Subsidence Monitoring in Tianjin Region

Land subsidence is a well-recognized environmental problem in some provinces in China, for example in the city of Tianjin in provinces Hebei and Jiangsu. The principal cause of the subsidence is continuous and excessive extraction of groundwater to fulfil the demands of fast development of industry and agriculture in both regions. For urban development and protection, it is necessary to monitor and measure the subsidence with very high vertical accuracy and spatial continuity at regular time intervals. Tianjin city in province Hebei and Wuxi, Suzhou, Changzhou region are chosen as the working area. The Tianjin municipality is situated on the coast of China on 117°15′ east, 39°10′ north about 200 km southeast of Beijing. It covers an area of about 3,000 km^2. Tianjin is a major Chinese city ("city province") with 9 million inhabitants. It is the main harbour for Beijing and is in the north of China. After the Shanghai and Beijing municipalities, Tianjin is economically the third most important centre of China. Subsidence due to groundwater extraction has become a major problem in this municipality. Governmental regulation has resulted in decreased subsidence. In the coastal areas outside Tianjin the subsidence rate is still high. The levelling data consist of point measurements with a density of approximately one point per 4 km^2. Continuous subsidence maps are produced by interpolation between the points. The collection of this information puts very high

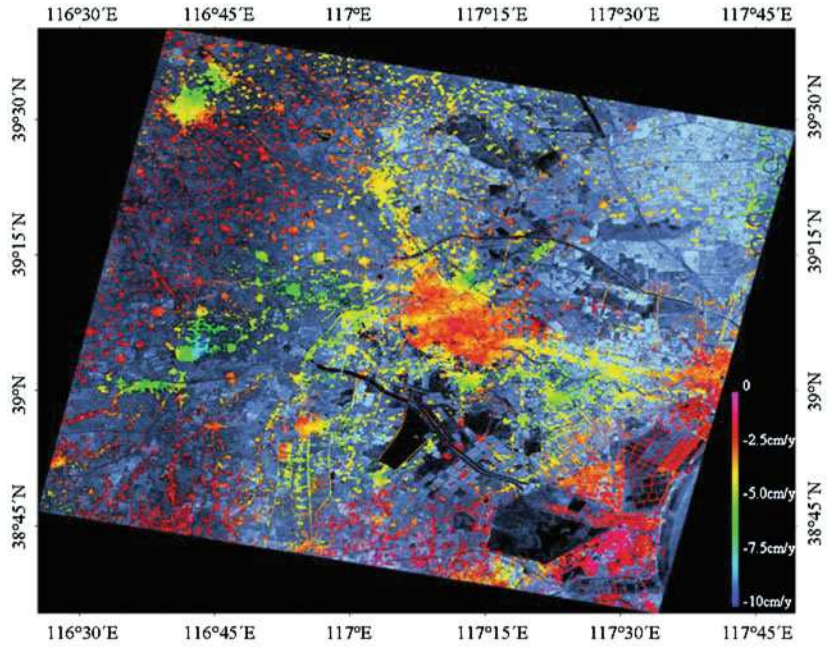


Fig. 11.22 Subsidence map of the Tianjin region (2005–2006)

demands on financial resources: in the order of US $1 million per year. For this reason the Tianjin Control Land Subsidence Office has actively been considering the potential of new technologies to improve the cost to benefit ratio of their work. One of these new techniques is the use of the Global Positioning System (GPS). In 1998 a GPS measuring network has been established in part of the area. Another relatively new technique with a high potential in this respect is SAR interferometry (INSAR). We used the ENVISAT ASAR data and D-INSAR technique to measure the subsidence in this region. Figure 11.22 shows the subsidence map of Tianjin from 2005 to 2006 derived from the differential SAR interferometry. Because the area to measure is relatively large, and only a few data sets (10) are provided for INSAR implementation, the conventional PS method is not suitable. Here the method used to estimate the average subsidence rate is still based on (53), but the PS candidates are substituted by coherent multi-looked points, which have relatively high coherence value, so that the phase stability can be ensured. The other advantage of using coherent multi-looked points is that the inaccuracy of the orbit parameters can be seen in each differential interferogram and therefore corrected and removed, while by conventional PS method it is "blended".

11.5 SAR Interferometry with Corner Reflectors (CR-INSAR)

The conventional differential INSAR technique has the potential for monitoring centimetre-scale ground motion in an accurate and cost-effective manner. The most important limiting factor in the application of INSAR is temporal change in the complex reflectivity of the ground surface during the period between radar acquisitions. This can be due to changes in parameters such as moisture content or vegetation. The persistent point scatterer techniques can take INSAR a step further by estimating atmospheric, orbital and DEM errors to derive very precise displacement and velocity measurements at specific points on the ground, but is limited only to the urban-like regions. The stable artificial corner reflectors can be identified from long temporal series of INSAR images even with large baselines and therefore decrease the risk of image decorrelation. In this section we represent the third INSAR analysis method: SAR interferometry based on the corner reflectors (Xia et al. 2002). This technique (CR-INSAR) is typically used for measuring and monitoring displacement of small-scale objects such as landslides and some specific civil engineering structures, e.g. water reservoirs, dams, bridges and buildings. We will discuss the following questions: (1) How to get the true location and phase of a corner reflector in an SAR complex image? (2) How to co-register the corner reflector pixels, if the coherence of its surrounding area is extremely low? (3) How to compute the interferometric phase of the corner reflectors without flat earth term and corners' height contribution? (4) How to estimate the CR's movement velocity? In order to demonstrate the results some practical examples of landslide monitoring in the Three Gorges area in China are given. The square trihedral offers an increased RCS (radar cross section) at the same beamwidth as the triangular trihedral, but is not as

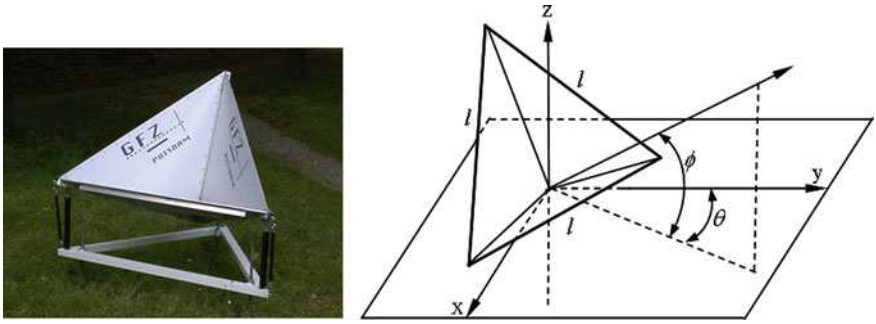


Fig. 11.23 Photo and schematic diagram of corner reflector

stable, especially in an environment where wind stress may produce plate bending. In this section we discuss the triangular trihedral corner reflectors only.

Figure 11.23 shows a schematic diagram of a corner reflector and the photo of a physical corner (made by the GeoForschungsZentrum, Potsdam) consisting of three flat triangular and perfect electric conductor planes orthogonal to each other. The corner reflector's actual radar cross section (RCS) is controlled by two angles ϕ and θ as well as the edge length l of the corner reflector aperture. The plane containing the line of radar sight and the crossing line from the corner reflector's common angle to peak point c is defined as the depression plane, and the plane perpendicular to the depression plane containing the radar sight line is called the datum plane. The angle between the radar sight line and the project line of the symmetry axis of the corner reflector on the datum plane is the azimuth angle θ, and the angle between the symmetry axis and its projected line on the datum plane is called aspect angle ϕ. Along the direction of the symmetry axis ($\theta = 0$, $\phi = 0$, namely the radar sight line and the symmetry axis over each other) the radar cross section (or backscattering coefficient) of the corner reflector reaches its maximum value (Uraby 1987):

$$\sigma_{\max} = \frac{4\pi l^4}{3\lambda^2}. \tag{55}$$

Generally, the radar cross section value will decrease if both angles θ and ϕ are not equal to zero at the same time. Because the 3-dB beamwidths in the θ plane (if $\phi = 0$) and in the ϕ plane (if $\theta = 0$) are the same, about 30°, the corner reflectors are very suitable for INSAR application. For ENVISAT satellite ASAR with wavelength $\lambda = 5.63$ cm and a small corner reflector with an edge of 1 m, the maximum radar cross section of the corner reflector $\sigma_{\max}$= 30 dB, which is easy to identify against its vegetative surroundings. Usually, in order to recognize a corner reflector in an SAR image, the difference of the backscattering coefficients of corner reflector and its surrounding should be larger than 20 dB; otherwise larger corner reflectors have to be used.

11.5.1 Orientation of the Corner Reflectors

As mentioned above, the 3-dB beamwidth of the aspect and azimuth angles is about 30°; this means that the radar can easily capture the corner reflector in a larger scope. However, orientation of the reflectors relative to radar has still to be carefully determined and measured in order to reach the maximum radar cross section, and guarantee that the radar sight line and the symmetry axis overlap with each other (i.e. $\phi = 0$ and $\theta = 0$). Therefore, the tilt angle of the corner reflector has to be adjusted relative to satellite's incidence angle (for example 19–26° by ERS and ENVISAT swath IS2). The azimuth angle of the radar sight line ϑ depends on the inclination of the orbital plane α and the geodetic latitude of the corner reflector ζ (Curlander and Robert 1991) and can be shown to be approximately

$$\vartheta = arc\,sin\left(\frac{\cos\,\alpha}{\cos\,\zeta}\right). \tag{56}$$

The approximation in (56) is strictly valid only for nadir pointing instruments and the location of the installed corner reflectors should be not too far from the equator (latitude < 60°). A more accurate approach to derive the azimuth angle ϑ is to use the location algorithm to determine the geocentric location of the satellite. It is known that the azimuth and aspect angles of the radar sight line may be derived from its direction cosine vector, i.e. the line connecting satellite and the corner reflector. Later, in Sect. 11.5.4 we will discuss how to get the location vectors of the corner reflectors and satellite where the corner reflector has just been acquired by radar.

11.5.2 Interpolation Kernel Design and Co-registration

Rate changes are useful in many applications, such as interconnecting digital processing systems operating at different rates and image co-registration. Sampling rate increase is accomplished by interpolation, that is, the process of inserting additional samples between the original low-rate samples. The inserted or interpolated samples are calculated by a so-called oversampling digital filter. Usually that is a FIR digital filter, because it operates at a fast rate. With respect to the fast timescale, the low-rate samples may be thought of as being separated by several zero samples. The job of the FIR filter, or interpolation kernel, is to replace the zeros by the calculated interpolated values. If a new sampling rate of L times higher than the low rate f_s is required, the corresponding FIR filter is called an L-fold interpolator and operates at $f_s' = Lf_s$. For the low-pass signal, the ideal L-fold interpolator filter should be a low-pass filter, operating at the fast rate f_s', with cutoff frequency equal to the low-rate Nyquist frequency $f_c = f_s/2 = f_s'/2L$, if expressed in units of the digital frequency, then $\omega' = 2\pi f/f_s'$, $\omega_s' = 2\pi$ and $\omega_c' = \pi/L$, and the passband gain is taken to be L instead of unity. The ideal impulse response coefficients are easily obtained from the inverse Fourier transform

$$d(k') = \frac{1}{2\pi} \int_{-\pi/L}^{\pi/L} L\, \mathrm{e}^{j\omega' k'}\, d\omega' = \mathrm{sinc}\left(\frac{\pi k'}{L}\right) \tag{57}$$

Let $x(n)$ denote the low-rate input samples, and let $x_{\mathrm{up}}(n')$ be high-rate or upsampled input samples consisting of the low-rate samples separated by L–1 zeros. With respect to the high-rate time index n', the low-rate samples occur once every L high-rate ones, that is, at integral multiples of L:

$$x_{up}(\mathrm{n}L) = x(n)$$

or

$$x_{\mathrm{up}}(n') = \begin{cases} x(n), & \text{if} \quad n' = nL \\ 0, & \text{otherwise} \end{cases}. \tag{58}$$

The output of the ideal FIR interpolator is obtained by the convolution of the upsampled input samples $x_{\mathrm{up}}(n')$ with the impulse response $d(k')$:

$$y(n') = \sum d(k') x_{\mathrm{up}}(n' - k'). \tag{59}$$

The ideal interpolator (57) and processes (58) and (59) ensure that the spectra of the input and output samples in passband are identical, but with different length of zero samples. The interpolation (59) can be done by zero-padding the Fourier transformation of the input signal followed by an inverse Fourier transformation to convert data back to spatial or time domain.

Because the corner reflectors are used in the case of very low coherence, the co-registration process of corners by means of searching the maximum coherency value with a moving window is impossible. In order to overcome this difficulty, a new method is suggested, namely searching the maximal amplitude value for each corner by means of interpolation in each image. For example, each piece of 32 × 32 pixel containing a corner in each image is interpolated using the interpolator mentioned above and the upsampling factor is taken to 16. After the interpolation, a process searching maximal amplitude is accomplished to determine the location of the corner with an accuracy of 1/16 pixel. In this case, it is not needed to co-register all complex SAR images to a unique master, but all the corners are one-to-one co-registered. The phase of each corner in each image is then extracted individually. If all corners are one-to-one co-registered and their phases are extracted individually, the next step is to calculate the phase pattern of flat terrain for each corner in an interferometric data pair.

11.5.3 Phase Pattern of Flat Terrain

The mathematical representations of the orbits (position and velocity) for the master and the slave satellites are identical, but with different coefficients:

$$\begin{cases} x_{(t)} = \sum_{n=0}^{4} a_n t^n \\ y_{(t)} = \sum_{n=0}^{4} b_n t^n \\ z_{(t)} = \sum_{n=0}^{4} c_n t^n \end{cases} . \tag{60}$$

The coefficients are derived from the precision orbit ephemeredes, such as one produced by GFZ/D-PAF in accordance with the image acquisition time for ERS (Reigber et al. 1996; Xia 2001) and ASAR Doris Orbit product from ESA for ENVISAT.

Starting from the acquisition time t_0 for the first pixel at the first azimuth line in the master image and (60) as well as PRF (pulse repetition frequency), the master satellite's position and velocity corresponding to each azimuth line can be determined. By means of the orbital state (position vector $\vec{S}$ and velocity vector $\vec{V}_s$) and using equations from Curlander (Curlander and Robert 1991), the pixel coordinate vector $P_{(x_p,y_p,z_p)}$ in an SAR image should satisfy the following non-linear equation system:

$$\begin{cases} \dfrac{\left(x_p^2 + y_p^2\right)}{R_e^2} + \dfrac{z_p^2}{R_p^2} = 1 & \text{ellipsoid} \\ \dfrac{2}{\lambda R_s}(\vec{V}_s - \vec{V}_p) \times (\vec{S} - \vec{P}) = f_d & \text{doppler} \\ \sqrt{(\vec{S} - \vec{P}) \times (\vec{S} - \vec{P})} = R_s & \text{distance} \end{cases} , \tag{61}$$

where R_e, and R_p are the radius of the earth at the equator and polar, respectively, $\vec{S}$ and $\vec{V}_s$ are the position and velocity vector of the master satellite, respectively, $\vec{P}$ is the position vector of target P, $\vec{V}_p$ is the velocity vector of target P, which depends on the target's position and the earth's rotational velocity, namely $\vec{V} = \bar{\omega} x \vec{P}$, f_d is Doppler centre frequency and R_s is the distance between the target P and orbit and can be determined from the slant range time.

Suppose that the target P lies on the perpendicular line to the slave orbit, the position of the slave satellite S' related to the object P can then be obtained by solving an equation indicating the minimal distance between the target P and the slave orbit.

Now we have the position vector $\vec{P}(x_p, y_p, z_p)$ of the target P, the master satellite position vector $\vec{S}\,(x_{(t)}, y_{(t)}, z_{(t)})$, and the corresponding slave satellite position vector $\vec{S}'\,(x'_{(t)}, y'_{(t)}, z'_{(t)})$. We can calculate the flat earth surface phase term for each corner with its position in an image for each line and pixel index:

$$\phi_{\text{flat}} = \frac{4\pi}{\lambda}(\overrightarrow{SP} - \overrightarrow{S'P}). \tag{62}$$

In this step, the three-dimensional location vectors $\vec{P}$, $\vec{S}$, and $\vec{S}'$ belonging to corner reflector and master as well as slave satellite, respectively, are known. Connecting corner reflector and satellite we get the vector $\overrightarrow{SP}_{(x_{(t)}-x_p, y_{(t)}-y_p, z_{(t)}-z_p)}$ and its direction cosine $\cos\alpha$, $\cos\beta$ and $\cos\gamma$; then the azimuth angle of the satellite sight line with respect to any corner reflector can be determined:

$$\vartheta = \tan^{-1}\frac{\cos\beta}{\cos\alpha}. \tag{63}$$

11.5.4 Elevation-Phase-Relation Matrix C_h and Phase Unwrapping

The element of the matrix C_h presents the relation between the interferometric phase of a corner and its elevation under a given interferometric geometry condition. Using this coefficient one can calculate the phase contribution of a corner due to its elevation in an interferogram. Because all the corners' coordinates and orbital vectors are known, we can easily determine the satellites' coordinates corresponding to each individual corner. In fact, the solution of the non-linear equation system (61) is the coordinate vector of a point target on the Earth's surface, whose elevation is zero. In order to find the elevation-phase relation we have to modify the first equation in (61) as follows:

$$h = \frac{\sqrt{x_p^2 + y_p^2}}{\cos\varphi} - \frac{R_e}{\sqrt{1 - \dfrac{R_e^2 - R_p^2}{R_e^2}\sin^2\varphi}}, \tag{64}$$

where

$$\varphi = \tan^{-1}\frac{z_p + \dfrac{R_e^2 - R_p^2}{R_p}\sin^3\theta}{\sqrt{x_p^2 + y_p^2} - \dfrac{R_e^2 - R_p^2}{R_e}\cos^3\theta}$$

and

$$\theta = \tan^{-1}\frac{z_p R_e}{\sqrt{x_p^2 + y_p^2} R_p}.$$

If we set $h = h_i$ and $h = h_i + 1$, respectively, for a certain corner reflector with the elevation h_i, substitute (64) to the first equation of (61), using (62) we can obtain the phase ϕ_h for $h = h_i$ and $h = h_i + 1$, respectively. The matrix C_h is then constructed from these phase differences for all the reflectors.

Corner reflectors are used only in the case that the coherence value is extremely low, and therefore the phase unwrapping of an interferogram is usually impossible. So far we have dealt with phase unwrapping of corner reflectors only on a sparse grid. Because all the corner reflectors are isolated points, the integration of phase difference can be done only from one corner to its nearest one with a priori information, that is, the elevation of all corners must be known first. In fact, this elevation data for all corners can be easily obtained by means of GPS surveying.

Let ϕ_i be the phase of the corner CR_i, and ϕ_j be the phase of the corner CR_j, which is the nearest neighbour of the corner CR_i. Considering the goal of the phase unwrapping procedure is to determine an integer k for all corners, k has to satisfy the following elevation difference equation:

$$C_j(\phi_j + 2k_j\pi) - C_i(\phi_i + 2k_i\pi) = h_j - h_i. \tag{65}$$

Here, h is height of corner reflector and C is height-phase coefficient, which indicates the height change per radian in an interferogram at a determined image coordinate. Each corner reflector has its own different C in different interferograms. The corner index i is starting at 0 meaning the reference point, and $k_0 = 0$ is assumed.

We calculate each k always from its nearest neighbour, in order to decrease the atmospheric effects.

11.5.5 Differential Interferogram Modelling

Suppose there are $M + 1$ SAR images and $N + 1$ corners. Select one image as master and one corner reflector as reference point. If possible, the corner reflector selected as a reference point should be installed in a previously known stable area. Starting from the $M + 1$ images and $N + 1$ corners, M interferograms consisting of the phase differences for all the $N + 1$ corners can be generated. Now the reference point's phase can be subtracted from the other N corners and can be expressed by the following equation:

$$\phi = C_l L^{\mathrm{T}} + C_p P^{\mathrm{T}} + C_h H^{\mathrm{T}} + \frac{4\pi}{\lambda} R + E, \tag{66}$$

where ϕ is interferometric phase matrix without flat surface contribution, containing M columns and N rows; each column belongs to a corresponding corner and each row belongs to a related interferogram; hence it is $N \times M$-dimensional; L is azimuth line index (referred to the reference point) vector for all the corners in co-registered image (M-dimensional); P is slant range pixel index vector (referred to the reference point) for all the corners in co-registered image (M-dimensional); H is relative elevation vector (referred to the reference point) for all the corners (M-dimensional); R is deformation matrix containing M columns and N rows; each column refers to a corresponding corner and each row refers to a related interferogram ($N \times M$-dimensional); C_l is slope vector, linear phase components along the azimuth line

(N-dimensional); C_p is slope vector, linear phase components along the slant range pixel direction (N-dimensional); C_h is elevation-phase-relation matrix containing the phase contributed to a unit elevation for each corner and each interferogram, and depending on orbital data and corners' geophysical coordinates as well as their elevation ($N \times M$-dimensional); λ is wavelength of the radar carrier, constant; and E is residue phase matrix including atmospheric effects and noise ($N \times M$-dimensional).

In (66), ϕ, L, P, C_h and H are given; the unknown R, displacements for all the reflectors in each interferogram, can be solved by minimizing the norm of the matrix E:

$$\underset{(C_l, C_p, R)}{MIN} \|E\|_{L^\infty}. \tag{67}$$

As soon as the phase unwrapping process is finished, the corresponding movement or movement rate of each corner reflector may be estimated directly from the unwrapped phase or by means of a phase regression analysis. In most cases, the differential phase difference between two corner reflectors is larger than π or less than $-\pi$, so that the phase unwrapping process is unavailable.

A more effective way to determine the CR's displacement is the estimation of the movement rate of the corner reflectors with the following modelling:

$$\phi_{im} = \frac{4\pi r_{im}}{\lambda} + \frac{4\pi B_m h_i}{\lambda R \sin\theta} + \phi_{im,\text{atmo}} + \phi_{im,\text{noise}}. \tag{68}$$

In (68), ϕ_{im} is the phase of the ith corner in the mth differential interferogram, r_{im} is displacement of the ith corner in the mth differential interferogram, B_m is the baseline, h_i is height error of the ith corner, θ is the incidence angle and $\phi_{im,\text{atom}}$ and $\phi_{im,\text{noise}}$ are the phases caused by atmospheric delay and noise, respectively. As mentioned above, the estimation of the displacement with the aid of phase unwrapping in space domain is very difficult, usually impossible. However, if the displacement for each corner reflector varies approximately linearly with time, in most cases the assumption holds, and (68) may be rewritten as

$$\phi_{im} = \frac{4\pi V_i T_m}{\lambda} + \frac{4\pi B_m h_i}{\lambda R \sin\theta} + \phi_{im,\text{atmo}} + \phi_{im,\text{noise}}. \tag{69}$$

Here V_i is the displacement rate of the ith corner and T_m is the time span of the mth interferogram. In addition, considering a corner reflector is installed at a previously known stable place, the phase of the reference corner can then be subtracted from (69). After subtracting the reference phase, each physical variable related to a corner in (69) now represents a relative one. Stacking the M interferograms will build N phase time series marked by ϕ_{ij}, $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, M$. Because the area of a landslide is usually not large, the distance of any two corners is not far apart, the atmospheric delay can be seen as same, so that (17) becomes as follows:

$$\phi_i = 2\pi f_{iv} t + 2\pi f_{ih} b + \phi_{i,\text{noise}}, \tag{70}$$

where $f_{iv} = 2V_i/\lambda$, $t = T_m$, $f_{ih} = 2h_i/\lambda R \sin\theta$, $b = B_m$ and $\phi_{i,\text{noise}}$ is an independent Gaussian noise with a zero mean. Equation (70) indicates that, for each corner reflector, its phase time series is a two-dimensional resonant signal with the domains time span t and baseline distribution b. Therefore the problem of the displacement estimation has been transferred to the problem of the spectrum estimation. There are many methods to solve the two-dimensional spectrum estimation problem, in which four important 2D spectral estimation methods are periodogram, autocorrelation method, covariance method and modified covariance method. In our case, the two domains, time span t and baseline distribution b, are usually irregularly sampled. In order to avoid the request of the phase unwrapping process, we select the periodogram method to estimate the frequency f_{iv} related to displacement rate and f_{ih} related to height error, if we have enough differential interferograms.

Based on (70), for each corner reflector we can construct a periodogram:

$$E_{(f_v, f_h)} = \left| \frac{1}{M} \sum_{m=1}^{M} \left(e^{j2\pi f_v t_m + j2\pi f_h b_m - j\phi_m}\right) \right|^2 . \tag{71}$$

Obviously, the amplitude of $E(f_v, f_h)$ will reach its maximum value 1 when both frequency variables approach their real value that suggests that the estimation of the frequencies (f_v, f_h) can be done by maximizing (71).

11.5.6 CR-INSAR Example: Landslide Monitoring in Three Gorges Area

As the biggest water conservancy project in China, the Three Gorges Project attracts more and more attention of the world. The reservoir is of a canyon and a river-like reservoir with a total length of about 600 km and average width of 1.1 km. The storage capacity of the reservoir reaches 39.3 billion m^3 with the normal pool level at 175 m. The length of the reservoir bank, the reservoir capacity, the number of resettlement and ecological influence are matchless up to now. A great number of the geological hazards such as landslide and rock-fall occurred along the banks of the Yangtze River, and many of them are in an active stage at present. The variations of the geological, ecological and atmospheric environments, especially the landslide and rock-fall phenomena in the Three Gorges area after the storage, have to be investigated. Three landslides, Xintan, Shuping and Kaziwan in the Three Gorges area in China, are chosen as the working area; see their radar image in Fig. 11.24.

A great landslide occurred at Xintan Town, Zigui County, Hubei Province, on 12 June 1985 (Liu 1988). Xintan landslide was 1,900 m long south-northwards and 210–710 m wide east-westwards. The landslide volume was about 30,000,000 m^3. The lands and stone that fell into the river were about 2,600,000 m^3. The tail skirt of the landslide and the boundary of the west part were bedrocks cliff composed of Devonian Permian sandstones and lime rocks. The east boundary was the fissure surface cut by the accumulation layers of the slope; here the accretions were

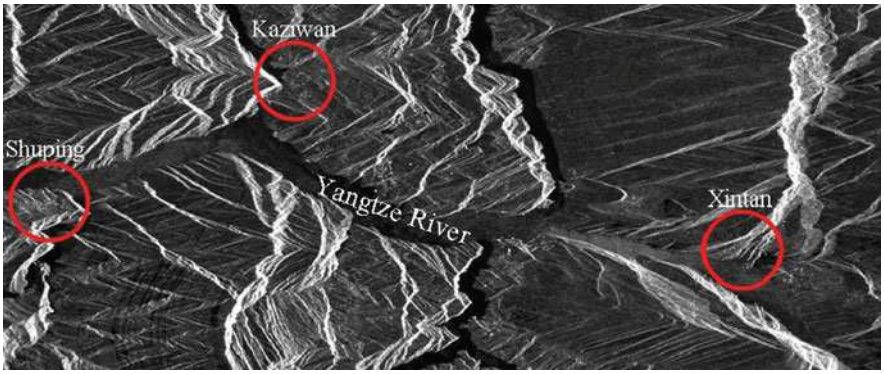


Fig. 11.24 Locations of Xintan, Shuping and Kaziwan landslides and installed reflector corners. *Left*: Shuping, 12 corners; *middle*: Kaziwan, 9 corners; *right*: Xintan, 4 corners. Acquisition time: 2003–2007; data: ENVISAT ASAR

30–40 m thick; some were 110 m thick. The residual body of the landslide passed through adjustment deformation for a few years, the lands and stones were fixed with time, subsidence and displacement decreased, and it seems to have a tendency of stabilizing. But the normal pool level will be 140–175 m, and one needs to know whether the stability of the residual landslide body can hold. Shuping and Kaziwan are another two active landslides not far from Xintan. The aim of choosing the three landslides as the working area is to verify the ability of the CR-INSAR technique that can measure not only the stable but also the moving landslides with higher accuracy and lower costs in large areas.

In order to monitor the stability of the Xintan landslide and the movement of the Shuping and Kaziwan landslides, 4 reflector corners were installed on the Xintan landslide in 2000, 9 corners in Kaziwan in August 2005 and 12 corners in Shuping in September 2005, respectively. Because of the cliff in the west boundary of Xintan landslide, the descending satellite orbit has been chosen and all the corners have been oriented to face the southeast according to (56). The photos of the three landslides and some corner reflectors are shown in Fig. 11.25. Each corner was then positioned by means of GPS measurement. Therefore the three-dimensional coordinates for each corner are known and used for computing its differential interferometric phase.

Up to now we have received more than 30 ENVISAT single look complex data sets delivered by ESA. In some images several corners disappeared because of unknown reasons; in addition, the corners on Shuping and Kaziwan landslides were installed later than on Xintan, and therefore not all the data can be available for the CR-INSAR studying in the area. Figure 11.26 is the radar intensity image of Xintan landslide, the corner 4 installed on a stable area was selected as reference, the results of another three corners from almost 4 years monitoring (see Fig. 11.27) indicating that the landslide at Xintan is very stable, in agreement with other field measurement result from GPS, conventional levelling, and trigonometric levelling in total station instruments.

Fig. 11.25 Photos of the landslides and reflector corners

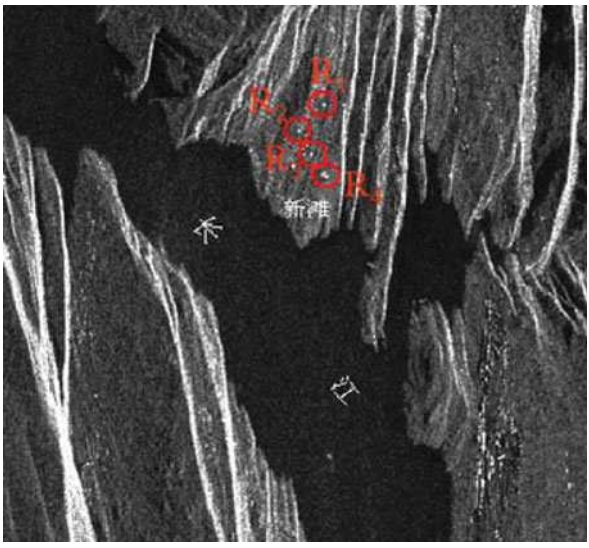


Fig. 11.26 Radar intensity of the landslide at Xintan and four installed corners

Figure 11.28 is the distribution map of the corner reflectors in Kaziwan landslide. Reflector R_2 and R_1 are installed on the stable area as reference and validation, respectively. The results of the INSAR measurement are drawn in Fig. 11.29. The result of R_1 shows it is really very stable. Corner R_3 and R_4 are located in front of

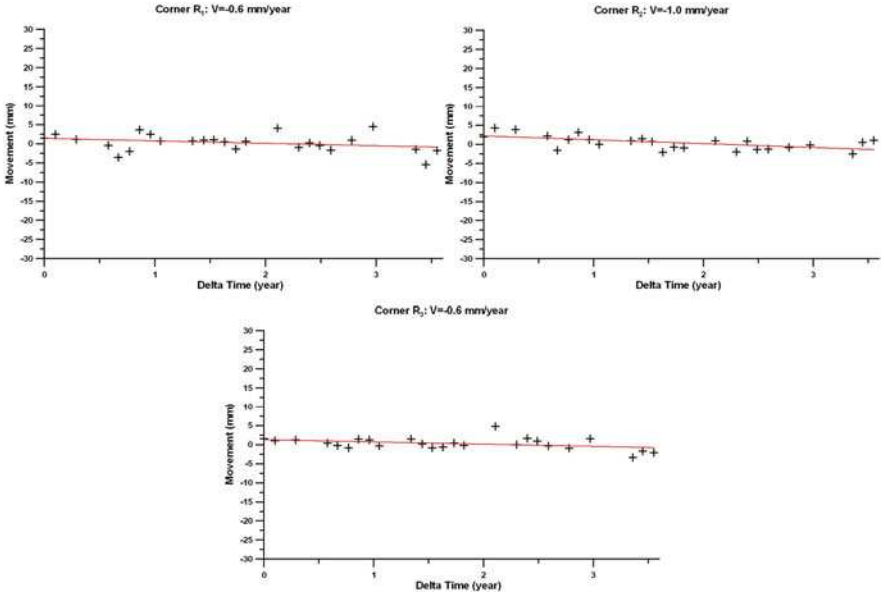


Fig. 11.27 Movements of the corners on the landslide at Xintan (20030817/20070617)

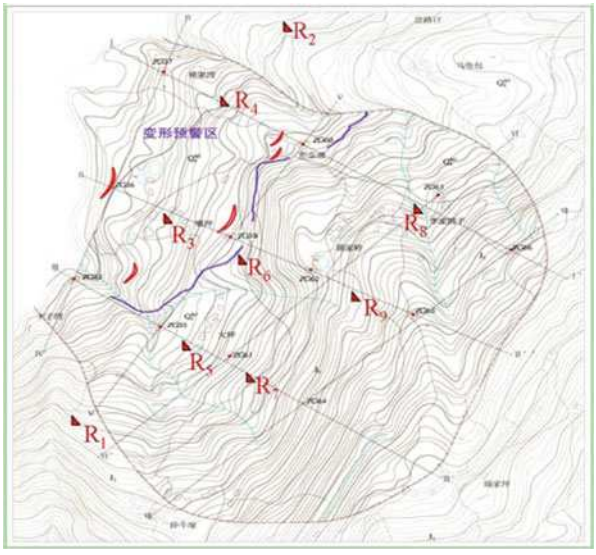


Fig. 11.28 Landslide at Kaziwan and corner distribution

the landslide faced to the river. As expected, they have large movements, while R_7, R_8 and R_9 move slowly.

In Shuping landslide 12 corners are installed, see Fig. 11.30. In fact, there are two sub-landslides here together, noted as block #1 and block #2; R_6 and R_8 are installed on the stable area, relative to reference corner R_8, R_6 shows stable, landslide #2 has

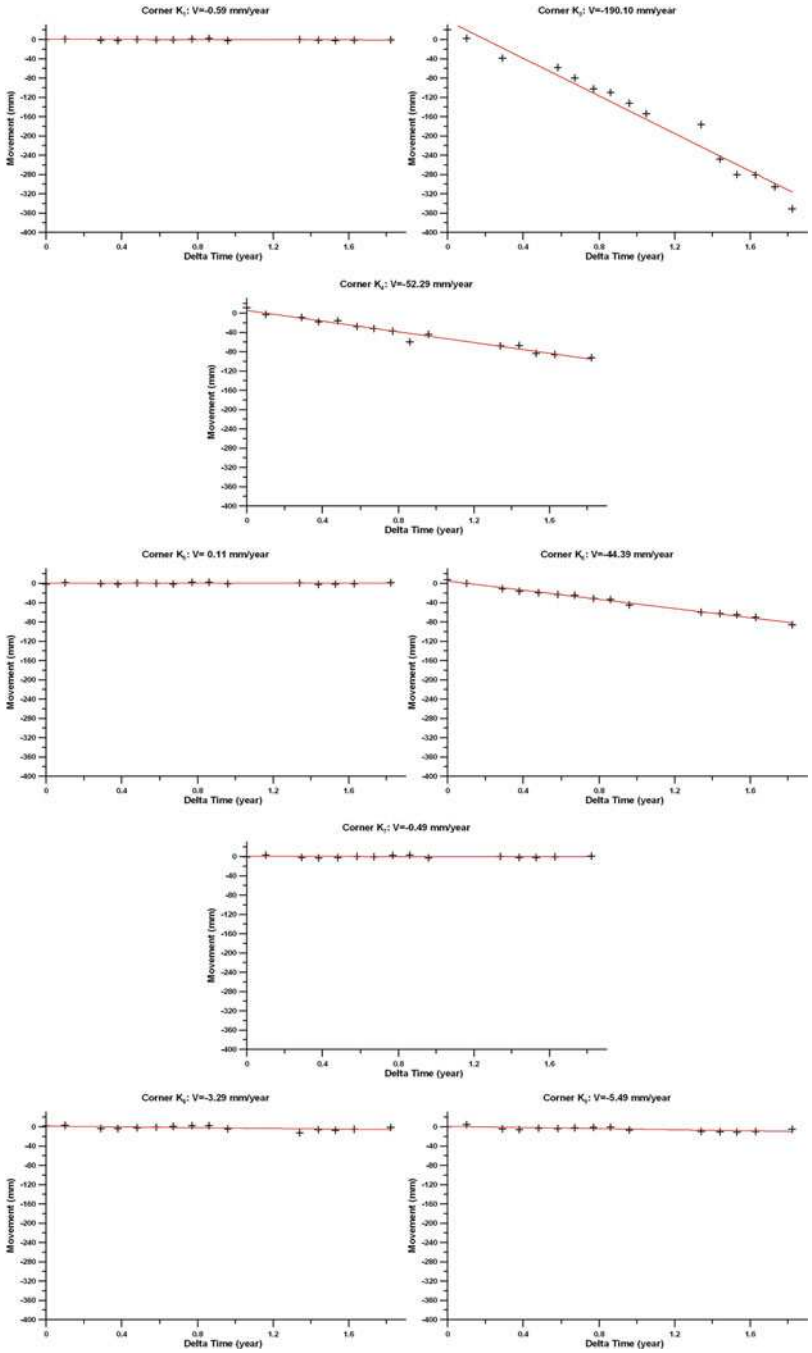


Fig. 11.29 Movements of the corners on the landslide at Kaziwan (20050821–20070617). Both R_2 and R_1 are installed on the stable area as reference and validation, respectively. The result of R_1 shows it is really very stable. Corner R_3 and R_4 are located in front of the landslide facing the river and, as expected, they have large movements, while R_7, R_8 and R_9, that stay, are moving slowly

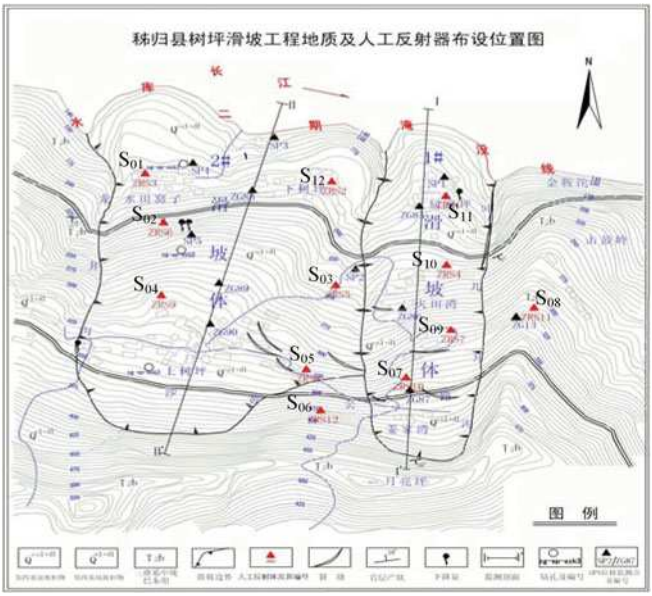


Fig. 11.30 Landslide at Shuping and corner distribution

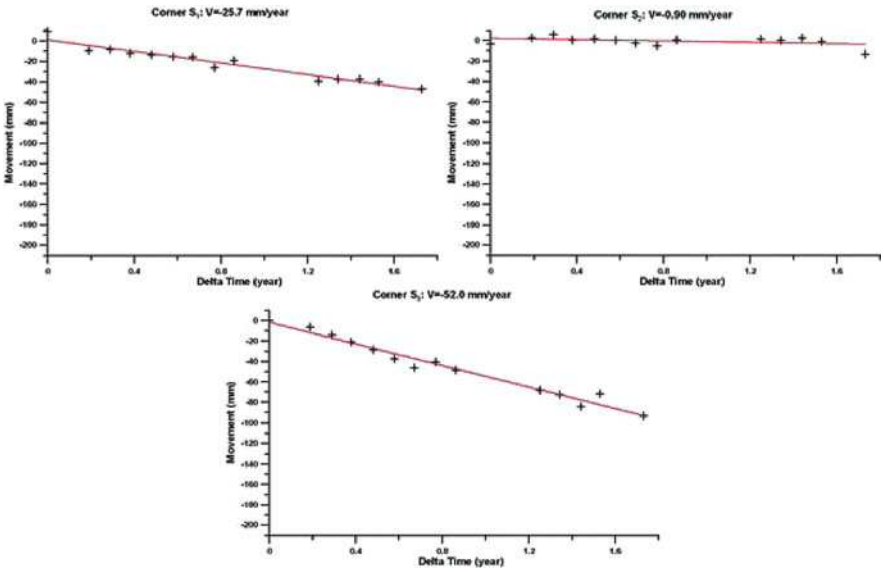


Fig. 11.31a Movements of the corners on the landslide at Shuping (20050925/20070617)

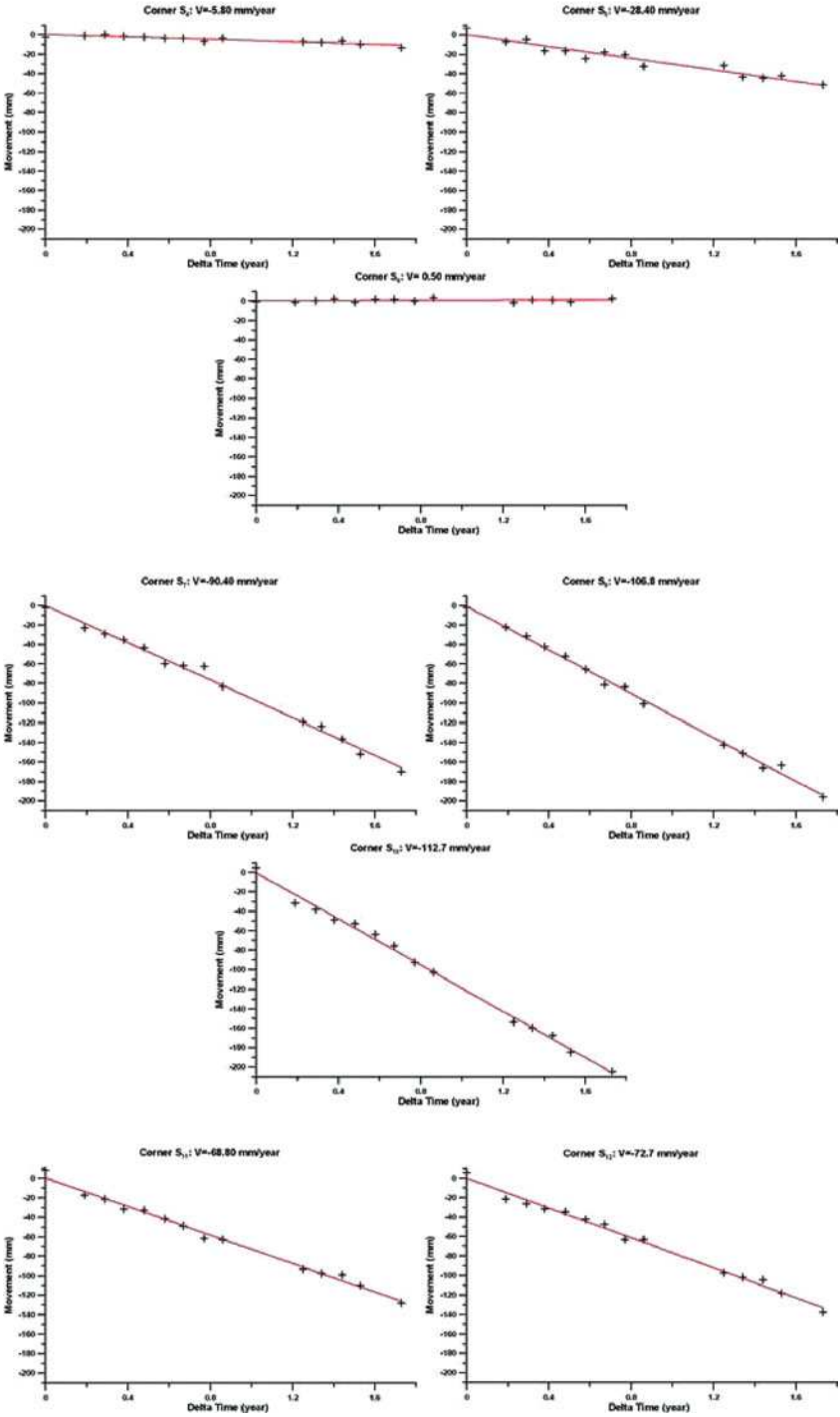


Fig. 11.31b (continued)

small displacement, 1–3 cm/year, while landslide #1 is moving relatively quickly. It is interesting that the results of the CR-INSAR here indicate that the movement of R_{11}, in the front of the landslide #1, is quite large, but the R_7, R_9 and R_{10} in the back of the landslide slide more quickly. The results of all the corners, indicated in Fig. 11.31a, b, coincide with the results by other local field measurements.

11.6 High-Resolution TerraSAR-X

TerraSAR-X is Germany's first national remote sensing satellite to be implemented in a public–private partnership between the German Aerospace Centre (DLR) and EADS Astrium GmbH. TerraSAR-X was launched on 15 June 2007 and went into operation on 7 Jan. 2008. DLR owns and operates the satellite and the payload ground segment (PGS) and holds the rights for the scientific exploitation of the data. Company Astrium holds the exclusive rights for the commercial exploitation of the data products. TerraSAR-X operates in three different modes: strip mode, spotlight mode and scan mode, with different resolution and coverage. For the aim to monitor landslide, the strip-mode data with 3-m resolution and 30-km image width have been selected. Spotlight-mode data have 1-m resolution, but the coverage is quite small and the cost is too expensive, while scan-mode data have large coverage, but lower resolution, so that the landslides in the image occupy only very few pixels. Some important system and orbit parameters are summed up in Table 11.2. Compared to ERS and ENVISAT, the TerraSAT-X data are suitable for landslide monitoring due to its high resolution (3m) and short repeat acquisition cycle (11 days). In most cases, the dimension of a landslide is several hundred metres long and wide, so that a landslide in an ERS or ENVISAT image covers only a few pixels after multi-look average processing. TerraSAR-X has higher resolution against the previous sensors; its high resolution benefits from the increase in radar bandwidth. Such a high level of image resolution and wide bandwidth make the traditional INSAR

Table 11.2 System and orbit parameters of TerraSAR-X

System and orbit parameters	
Radar carrier frequency	9.65 GHz
Wavelength	3.1067 cm
Nominal orbit height at the equator	514 km
Incidence angle range	20–45° (full performation) (15–60°accessible)
Polarizations	HH, VV, HV, VH
Antenna length	4.8 m
Nominal look direction	Right
Antenna width	0.7 m
Pulse repetition frequency (PRF)	2.0–6.5 kHz
Chirp bandwidth	150 MHz max. (300 MHz experimental)
Resolution	1 m (spotlight), 3 m (strip) and 16 m (scan)
Orbits/day	$15^{2/11}$

method to be applied for monitoring landslide, even with large baseline of over 3,000 m. This limit can be derived by use of (72), (73) and radar system parameters from Table 11.2.

It is known that the fringe frequency in slant range (the derivative of the interferometric phase to slant range distance) can be represented as a function of perpendicular baseline $B_\perp$, wavelength λ, slant range R and incidence angle θ:

$$f_{\mathrm{r}} = \frac{2B_\perp}{\lambda R \tan(\theta)}. \tag{72}$$

This fringe frequency indicates a spectral frequency shift between the two acquisitions caused by orbit offset, namely so-called baseline. The critical baseline will be determined by the maximum allowable fringe spectral shift $f_{r,\max}$, which depends on the wave velocity c and radar chirp bandwidth B (inverse of the pulse duration):

$$f_{r,\max} = \frac{2B}{c}. \tag{73}$$

The critical baseline in the case of TerraSAR-X might be several thousands of metres derived from (72) and (73), while by ERS/ENVISAT C-band sensors it is only about 800 m. The traditional D-INSAR method can't be effectively applied to landslide monitoring, if only ERS or ENVISAT data are used (Xia et al. 2004). In the cases of ERS and ENVISAT, the repeat cycle is 35 days and the resolution is 30 m. The long repeat cycle and low resolution will cause the de-coherence phenomenon, especially in the vegetated Three Gorges area. On the one hand, by contrast, the second advantage of the TerraSAR-X data is their short repeat acquisition cycle. In a short time interval more data sets can be delivered for persistent targets analysis. On the other hand, maybe just some events happen in the 11-day cycle; if so, these events can be monitored on time with no major delays. The TerraSAR-X data are very suitable for deformation monitoring applications, such as landslide and subsidence, because of their short wavelength (increasing the measurement performance) and short repeat cycle (capturing the movements on time).

Up to now, 18 TerraSAR-X strip-mode single look complex images covering the working area have been delivered by DLR. The data acquisition interval is from 21 July 2008 to 4 Feb. 2009. As mentioned above, due to height resolution and short repeat acquisition cycle, the TerraSAR-X strip-mode data are very suitable for global monitoring in short intervals. Using the classic D-INSAR method, namely a two-pass differential interferometry approach, the stability or instability of the landslides in the considered area covered by SAR images may be continuously monitored in a short period. For global landslide monitoring, all the possible differential interferograms were calculated. Thanks to the TerraSAR-X operating control, the baselines of all the possible interferometric data pairs are quite small, less than 450 m. Hereby SRTM 3-arc-second DEM was used as height reference. In order to demonstrate the results of the classic D-INSAR technique by use of TerraSAR-X strip-mode data for landslide monitoring in the area, six differential interferograms are selected and shown in Figs. 11.32, 11.33, 11.34, 11.35, 11.36 and 11.37 to validate the effectiveness of the TerraSAR-X interferometry for landslide

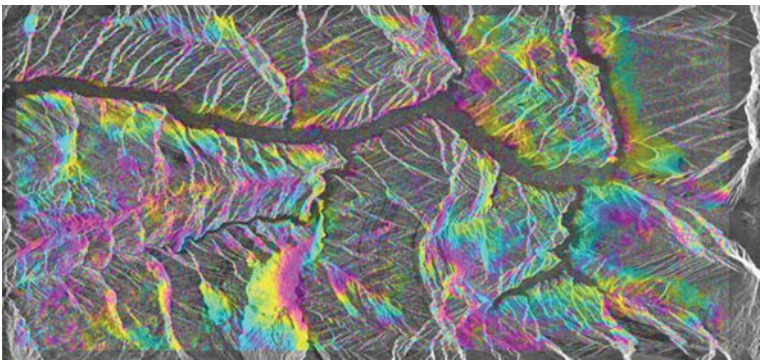


Fig. 11.32 Differential interferogram of data pair 20080721/20080914; time interval is 55 days and baseline is 10 m. Colour cycle from *red* via *yellow* to *blue* represents a displacement of 1.5 cm

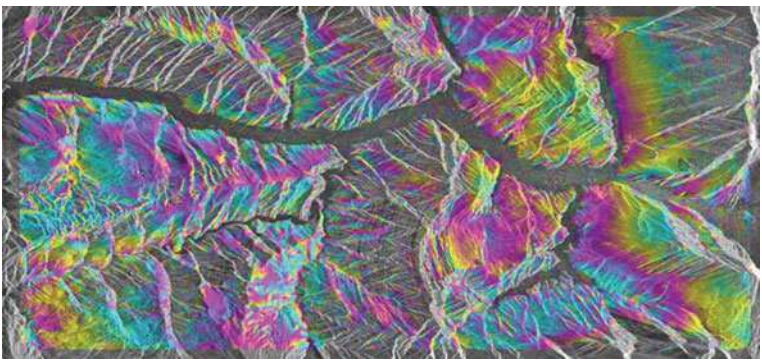


Fig. 11.33 Differential interferogram of data pair 20080914/20081017; time interval is 33 days and baseline is 20 m

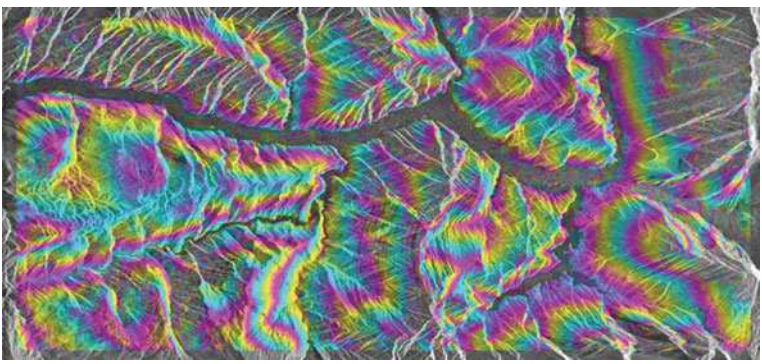


Fig. 11.34 Differential interferogram of data pair 20081017/20081119; time interval is 33 days and baseline is 111 m

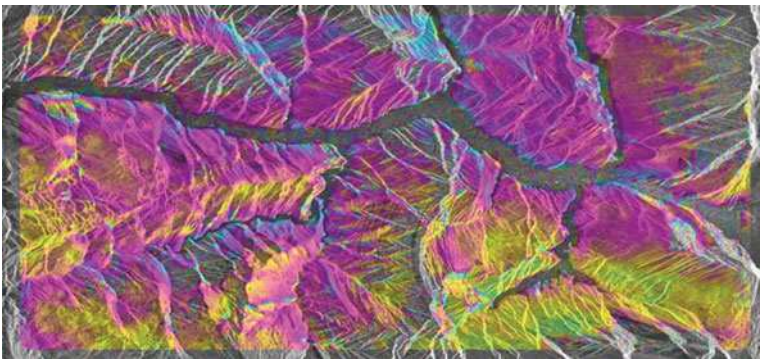


Fig. 11.35 Differential interferogram of data pair 20081119/20090113; time interval is 55 days and baseline is 67 m

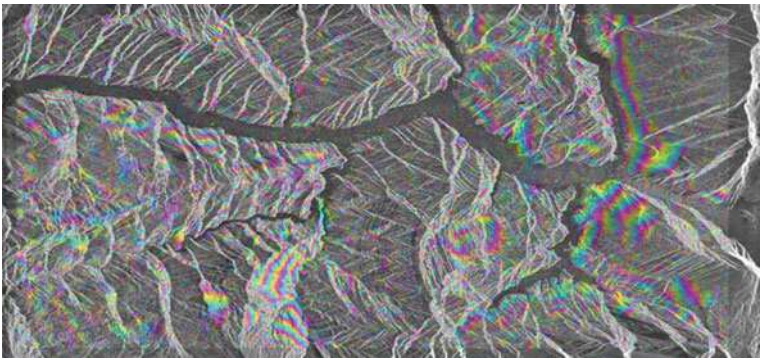


Fig. 11.36 Differential interferogram of data pair 20080721/20090113; time interval is 176 days and baseline is 13 m. Clearly, the displacement in this interferogram is a sum of the displacements in the former four interferograms. The interferogram series shows that in the 176-day monitoring interval, the displacement of the mountain slope happened mainly from September to mid-November, just corresponding to the water filling period. Each colour cycle indicates 2π phase difference, meaning a displacement difference of 1.5 cm. The increasing direction is from *red* via *yellow* and *green* to *blue*

monitoring and show the displacements derived from this technique or instability of the whole working area. These six differential interferograms are organized in chronological order to show the displacement history. According to traditional practice, the interferometric phase is represented by colour, each colour cycle meaning a phase difference of 2π, which corresponds to a displacement of 1.55 cm in line-of-sight. Figure 11.32 is the differential interferogram of data pair 20080721/20080914 with a baseline of 10 m. In the 55 days there were small displacements to see. Figure 11.33 is the differential interferogram from data pair 20080914/20081017 with a baseline of 20 m. In this 33-day time interval, large movement happened almost everywhere, the most obvious being in the landslide at Fanjiaping, Xintan,

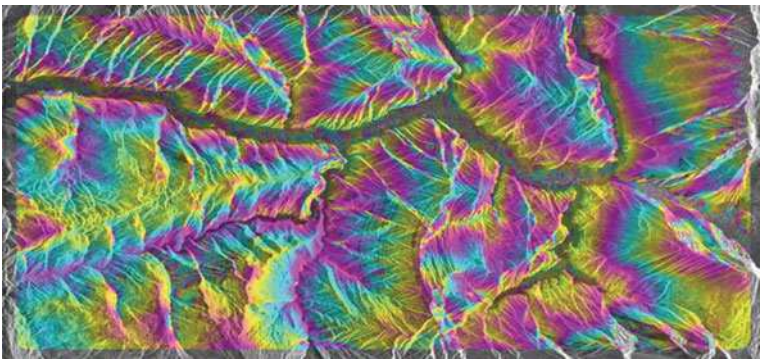


Fig. 11.37 Differential interferogram of data pair 20090102/20090204; time interval is 33 days and baseline is 47 m. On and after 2 Nov. 2008 the water level began to decrease. In the downing period, compared to filling period, the direction of the movement grad is inverted. From the top to the foot of a mountain the colour cycle rank is from *red* via *yellow*, *green* to *blue*, which means the displacement at mountain foot is larger than the displacement on mountain top

and on some mountain slopes, such as the left bank of Xiangxi river and the slope of the opposite mountain. The land surface displacement of this area in the time span of 20081017/20081119 – see Fig. 11.34 – is like that in Fig. 11.33 with same time interval, but large baseline of 111 m. Figure 11.35 is the differential interferogram of data pair 20081119/20090113 with a baseline of 67 m. Unlike the former three interferograms, this interferogram shows that in the 55-day interval almost no movement happened except the Fanjiaping landslide. Figure 11.36 is the differential interferogram of data pair 20080721/20090113 with a baseline of 13 m. Clearly, the displacement in this interval is the sum of the former four interferograms. In fact, the Three Gorges Project is a multi-functional water control system. Its key functions include flood control and power generation. The project has been constructed in three phases. Storing water at the 175-m level is a requirement for the last phase of the dam construction. From September to November, the time interval of the data acquisitions, the State Council's Three Gorges Project Construction Committee raised the water level to 175 m to test the storing function that began in late September. The committee expects the higher water level would not cause massive silting, nor lead to serious natural disasters in nearby areas. The results of these TerraSAR-X interferometries show, however, that despite no serious natural disasters happening, potential risks are everywhere to be seen. In the interferogram of data pair 20080721/20090113, Fig. 11.36, one colour cycle on the landslide Xintan and several colour cycles on the landslides Fanjiaping indicate that the movements of the two landslides were increasing with time. Because the landslides at Shuping and Kaziwan are covered by vegetations and fruit trees, radar signal was not coherent, so there was no sufficient and clear interferometric phase to see. Certainly, the differential interferometric phase contains not only the contribution of the displacements, but also the contribution of the orbit inaccuracy. According to INSAR theory, the phase contribution caused by orbit error in these five interferograms should be

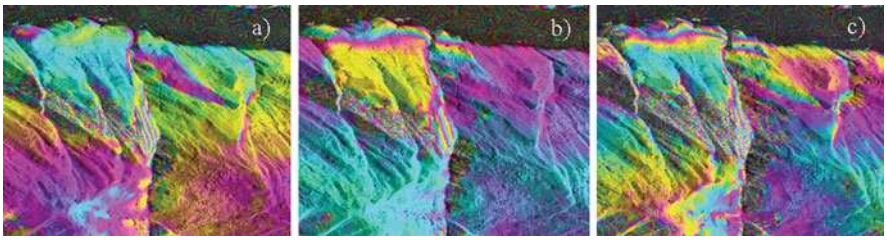


Fig. 11.38 Differential interferograms of landslide at Fanjiaping derived from a data pair series of 20090102/20090113 (**a**), 20090102/20090124 (**b**) and 20090102/20090204 (**c**) showing the landslide mass movement history and accumulation

excluded, because the interferometric phase or fringes are not correlated with the corresponding baselines. Therefore, the contribution of the displacements must be dominant in the differential interferograms. Figure 11.37 is a differential interferogram in the water level downing period, showing that the displacement of the mountain slopes was increasing from mountain top to mountain foot, just inverted if compared to the water level filling period. Figure 11.38 is a differential interferogram series of a region with a size of about 3×4 km^2 and 11-, 22- and 33-day intervals, respectively. Landslide Fanjiaping is located in the area. The displacement history and distribution of the whole landslide are very clearly exhibited. The small displacement rate, but with larger gradient, took place in front of the landslide. There are many other different landslide survey methods and techniques, but none of them may produce such a detailed landslide movement map.

References

Colesanti, C., Ferretti, A., Novali, F., Prati, C. and Rocca, F. (2003) SAR monitoring of progressive and seasonal ground deformation using Permanent Scatterers technique: IEEE Trans. Geo. Remote Sens., 41, 1685–1701

Costantini, M. (1996). A phase unwrapping method based on network programming. Geosci. Remote Sens., 36(3), 813–821

Curlander, J.C. (1982) Location of spaceborne SAR imagery. IEEE Trans. Geo. Remote. Sens., GE-20(3), pp. 359–364

Curlander, I.C. and Robert, N.M. (1991) Synthetic Aperture Radar System and Signal Processing. Wiley & Sons, INC. ISBN 0-471-85770-X, Printed in the USA

ESA, ASAR Product Handbook, issue 2.2, 27 February 2007, http://envisat.esa.int/handbooks/asar/

Ferretti, A., Prati, C. and Rocca, F. (2001) Permanent scatterer INSAR interferometry. IEEE, 39, 8–20

Ghiglia D.C. and Romero L.A. (1989) Direct phase estimation from phase differences using fast elliptic partial differential equation solvers. Opt. Lett., 14, 1107–1109

Goldstein, R.M. and Zebker, H.A. (1987). Interferometric radar measurement of ocean current. Nature, 328(6132), 707–709

Goldstein, R.M., Zebker, H.A. and Werner, C.L. (1988) Satellite radar interferometry: two-dimensional phase unwrapping. Radio Sci., 23(4), 713–720

Graham, L. (1974). Synthetic interferometry radar for topographic mapping. Proc. IEEE, 62(6), 763–768

Hunt, B.R. (1974) Matrix formulation of the reconstruction of phase values from phase differences. J. Opt. Soc. Am., 69, 393–399

Jakowatz, C.V., Wahl, D.E., Eichel, P.H., Ghiglia, D.C. and Thompson, P.A. (1996) Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach. Springer, Albuquerque

Liu, G.G. (1988) Environment geologic investigation of Xintan landslide. Environ. Geol., 12(1), 11–13

Pritt, D.M. and Shipman, J.S. (1994) Least-squares two-dimensional phase unwrapping using FFT's". IEEE Trans. Geosci. Remote Sens., 32(3), 706–708

Reigber, Ch., Xia, Y., Kaufmann, H., Massmann, F.-H., Timmen, L., Bodechtel, J. and Frei, M. (1996) Impact of precise orbits on SAR interferometry. Proceedings of the *ESA*-Fringe Workshop on Applications of ERS SAR Interferometry, ESA SP-406, Zürich, el-pub, http://www.geo.unizh.ch/rsl/fringe96

Uraby, F.T. (1987). Microwave Remote Sensing, Vol. 2. Chinese Scientific Publishing, Beijing, p. 204

Wang, R., Xia, Y., Grosser, H., Wetzel, H.-U., Kaufmann, H. and Zschau, J. (2004). The 2003 Bam (SE Iran) earthquake: precise source parameters from satellite radar interferometry. Geophys. J. Int., 159, 917–922

Werner, C., Wegmuller, U., Strozzi, T. and Wiesmann, A. (2003) Interferometric Point Target Analysis for Deformation Mapping. IGARSS 2003, Toulouse, 21–25 July 2003, pp. 4362–4364

Xia, Y. (1996) Bestimmung geodynamischer Effekte mit Hilfe der radarinterferometrie. Dissertation, Deusche Geodätische Kommision bei der Bayerischen Akademie der Wissenschaften, München 1966, Heft Nr. 467, ISBN 3769695070, p. 17

Xia, Y. (2001) INSAR Activities in Central Asia Using Mobile SAR Receiving Station. IGARSS 2001, Sydney, pp. 407–409

Xia, Y. (2005) Bam earthquake: Surface deformation measurement using radar interferometry. Acta. Seismologica. Sinica., 18(4), 451–459

Xia, Y., Kaufmann, H. and Guo, X.F. (2002) Differential SAR Interferometry Using Corner Reflectors. IGARSS 2002, Toronto, 24–28 June 2002, pp. 1243–1246

Xia, Y., Kaufmann, H. and Guo, X.F. (2004) Landslide monitoring in the Three Gorges area using D-INSAR and corner reflectors. Photogramm. Eng. Remote Sens., 70(10), 1167–1172

Index

G. Xu (ed.), *Sciences of Geodesy – I*, DOI 10.1007/978-3-642-11741-1,