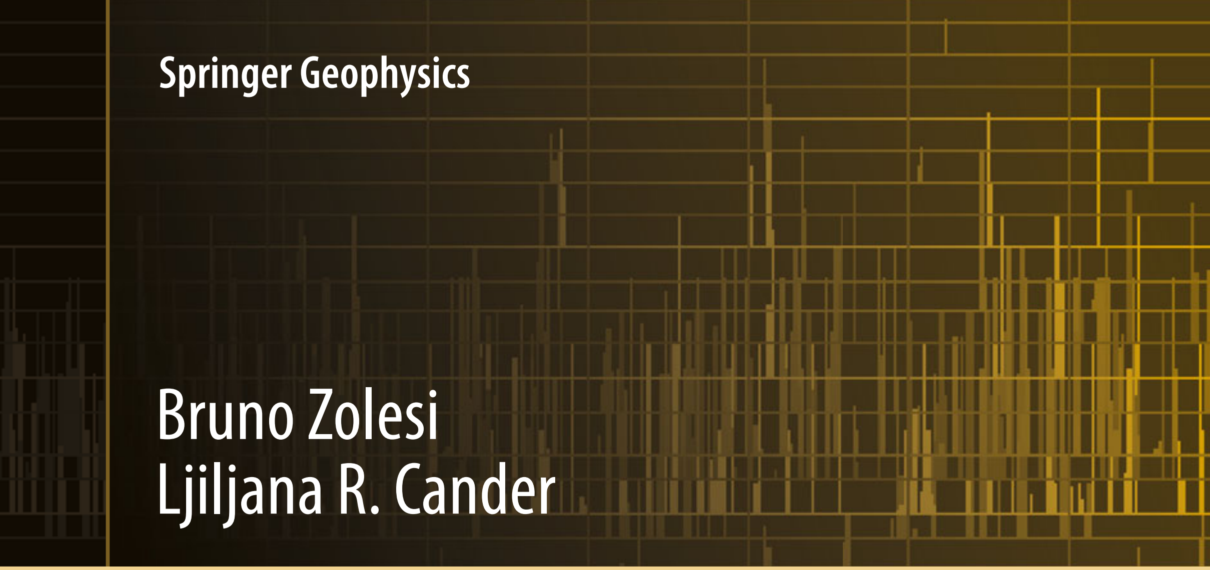


Springer Geophysics

Bruno Zolesi
Ljiljana R. Cander

# Ionospheric Prediction and Forecasting

Springer

# Ionospheric Prediction and Forecasting

# Springer Geophysics

For further volumes:
http://www.springer.com/series/10173

Bruno Zolesi · Ljiljana R. Cander

# Ionospheric Prediction and Forecasting

Springer

Bruno Zolesi
Istituto Nazionale di Geofisica e
Vulcanologia
Rome
Italy

Ljiljana R. Cander
Rutherford Appleton Laboratory
Harwell Oxford
UK

ISBN 978-3-642-38429-5 ISBN 978-3-642-38430-1 (eBook)
DOI 10.1007/978-3-642-38430-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013942473

© Springer-Verlag Berlin Heidelberg 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

*This book is dedicated to:*
*Paola and Giulia*

Bruno Zolesi

*Rudolf and Javorka*

Ljiljana R. Cander

# Acknowledgments

Our acknowledgements to the Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy, and the Rutherford Appleton Laboratory (RAL), United Kingdom, for supporting our research in ionospheric science and radio propagation over the years, thereby making our contributions to this work possible.

We have been extremely privileged to collaborate for many years with Prof. Pietro Dominici from the former Istituto Nazionale di Geofisica, who has left a valuable legacy for later generations to help improve mankind's knowledge of the Earth's ionosphere, and Mr. Peter A. Bradley from RAL, who created the first European COST 238 Prediction and Retrospective Ionospheric Modelling over Europe (PRIME) project, and has made profound contributions to experimental and theoretical studies in ionospheric prediction and forecasting.

Many thanks are due to our editors Dr. Chris Bendall and Ms. Helen Rachner, and all the people at Springer for doing such a wonderful job. We are indebted to Ms. Francesca Di Laura for her help with the Figures. Further thanks are due also to Mr. Gavin Taylor for his valuable English language support.

Finally, we would like to express our gratitude to our friends and colleagues Dr. Cesidio Bianchi, Dr. Luigi Ciraolo, Dr. Antonio Meloni, Dr. Michael Pezzopane, Dr. Alessandro Settimi, Prof. Sandro M. Radicella, Dr. Paolo Spalla, and Dr. Enrico Zuccheretti whose comments, suggestions, and stimulating discussions significantly improved the material presented in this book.

# Contents

# Chapter 1
# Introduction

## 1.1 Background

This book describes how to predict and forecast the state of the planet Earth's ionosphere under quiet and disturbed conditions, within achievable limits, in terms of dynamic processes in the weakly ionized plasma media of the upper atmosphere, using modern measurement and modelling techniques. In particular, it explains the close relationship between the state of the media, and radio wave propagation through this media. The prediction and forecasting algorithms, methods and models described take a practical approach to the design and engineering of ionospheric dependent systems. This means the definition of a system that permits investigation of ionospheric weather, which is the state of the ionized media at a given time and place, and ionospheric climatology, the average of ionospheric events over a long period. A proper understanding of the ionosphere is of fundamental practical importance for telecommunication and navigation systems that use the ionosphere to function, or that would function much better in its absence on the Earth or on any planet with an atmosphere.

Our aim when writing this book was to fill an obvious gap in contemporary literature, which lacks a coherent presentation of contemporary research directly related to ionospheric prediction and forecasting. Despite its great importance, such detailed ionospheric description has until recently maintained a rather low profile within the scientific and engineering communities. More theoretically oriented and largely experimental ionospheric research and descriptions are widely known. These aim at monitoring and understanding the nature of the terrestrial ionosphere in general, but very often do not include the sort of technical details currently utilized in ionospheric prediction and forecasting software. Recent advances in global and regional networks of real-time ionosondes and ground-based dual-frequency GNSS (Global Navigation Satellite System) receivers, together with data-driven ionospheric modelling, have brought new insights to ionospheric physics with significant improvements in prediction and forecast models. This book describes research and development that exploits the new potential to enhance

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_1,

understanding of how the Sun affects the Earth's atmosphere, helping operators and users to improve the efficiency and reliability of their terrestrial and space-based technological systems.

The authors have been active researchers in the ionospheric field for a number of years, and were very pleased to compile in a single volume their own research results, together with existing knowledge from their colleagues working in related areas, in order to provide a unified perspective. The intended audience of the present volume includes not only the ionospheric research community but also a broader range of scientists and engineers in related fields including atmospheric physics and climate, geodesy, geomagnetism and paleomagnetism, seismology, magnetospherics, solar and heliospheric physics, and aeronomy. It provides a valuable text for graduate, postgraduate, and post doc. students as well as staff at ionospheric stations and geomagnetic observatories around the world, and all those involved in applied science, military defence and security organizations, space and telecommunications companies and agencies. The authors hope that this volume will encourage members of the scientific community, engineers and students to undertake research in terrestrial and planetary ionospheres, stimulating scientific progress. Furthermore, used step-by-step, the present volume represents a concise operational handbook for forecasters of ionospheric climatology, for diverse weather and propagation services worldwide. References are provided whenever possible, with suggested reading after each chapter and a bibliography of the most significant books, to direct readers to the most useful and relevant literature on the subject.

## 1.2 Previous Studies

### *1.2.1 A Brief History of Ionospheric Measurement*

Since the beginning of the last century the evolution of ionospheric science has followed the development of radio propagation measurements and studies. While the existence of ionized layers in the upper atmosphere had been postulated during the eighteenth century by various scientists like Balfour Stewart and Arthur Schuster, it was only on 12th December 1901 that the first transoceanic radio link realized by Guglielmo Marconi provided experimental proof. G. Marconi and his assistant George Kamp received the 3 dots of Morse code representing the letter S from the 120 meter wire antenna located in St. John of Terranova, in Canada. The transmitter was a 10 kW device located in Poldhu, Cornwall, UK, more than 3,000 km away. Despite numerous discussions after and before this experiment regarding the technique, frequency, and equipment used, other subsequent radio links confirmed that long distance wireless communication was possible in spite of the hindrance caused by the curvature of the Earth. Guglielmo Marconi was awarded the Nobel Prize in 1909 for his work.

Within a few months, in 1902, two scientists Oliver Heaviside and Arthur E. Kennelly independently proposed the same explanation: the upper atmosphere is characterized by the presence of charged particles so that radio waves can be transmitted across the Atlantic Ocean not by line of sight but within a conducting cavity between this particular region and the Earth. At the time the region came to be known as the Kennelly Heaviside layer or layers and was investigated by numerous experimental scientists, revealing the morphology of this part of the planet. The vertical structure of the region was finally described during the nineteen twenties in the systematic experiments and theoretical studies of Edward V. Appleton, winning him the Nobel Prize in 1947. This was possible thanks to the technological developments of the Americans Gregory Breit and Merle A. Tuve. In the same period the name “ionosphere” was coined during a discussion between another British scientist Robert Watson Watt and Appleton. Very soon the term ionosphere was also adopted in other countries in Europe, marking the beginning of systematic ionospheric measurements.

Starting from about 1930 the network of ionospheric vertical sounding stations was expanded considerably and the increased data contributed to a better knowledge of ionospheric phenomena. The stations were equipped with different types of “ionosondes”, which applied basic radar techniques to detect the electron density of the ionospheric plasma as a function of height, by scanning for transmissions from 1 to 20 MHz and measuring the time delay of any echoes. The complete story of this instrument and its evolution during the last century from large scale installations distributed through 3 rooms, to the fundamental improvements resulting from digital technology, is recounted by one of the protagonists, *Klaus Bibl,* in Annals of Geophysics, 1998.

While the first routine ionospheric sounding stations were established primarily for scientific purposes, there was considerable expansion during the Second World War due to the need for predicting long distance radio links. At that time the geophysical parameters measured were a function of radio communication frequency planning, rather than an effort to understand ionospheric phenomena and the physical mechanisms involved. The maximum expansion of the ionospheric vertical sounding network was achieved during the International Geophysical Year (IGY, July 1957–December 1958) (Fig. 1.1).

These ground-based routine measurements constituted the basis for the global models of ionospheric plasma. They were so important for any future ionospheric investigations that some international organizations concerned with establishing internationally agreed global propagation models, like the International Union of Radio Science (URSI), the Committee on Space Research (COSPAR), and in particular the International Radio Consultative Committee (CCIR), the forerunner of the International Telecommunication Union—Radio-Communication Sector (ITU-R), a United Nations Specialised Agency, strongly encouraged this kind of ionospheric observation with international data exchange through the World Data Centers.

In addition to vertical sounding, other instrumental techniques were introduced at ground-based stations, including new types of radar at EISCAT (European

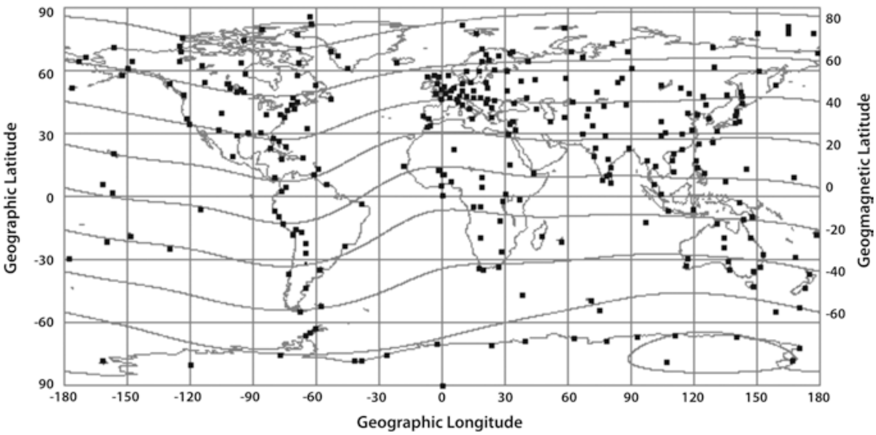


**Fig. 1.1** Map of the ionospheric vertical sounding stations operating during the International Geophysical Year

Incoherent Scatter) and SUPER-DARN (Super Dual Auroral Radar Network) facilities, as well as on-board rockets and satellites. It should be noted that satellites play a double role in ionospheric observations: both with measurements in-situ, and using the global navigation constellations as transmitting points towards the large number of GNSS receivers on land.

### *1.2.2 Evolution from Long-Term HF Planning to Nowcasting and Space Weather Applications*

Over the last four decades, models for long-term prediction of the median monthly conditions for HF (high frequency 3–30 MHz) radio propagation have represented an important tool not only for applied science and for radio users, especially frequency planners at radio broadcasting agencies, but also for geophysicists in their theoretical studies of the upper atmosphere. The systematic collection of regular observations and the evolution of computing systems allowed William B. Jones and Roger M. Gallet to produce, in 1965, the first significant global model for long-term mapping of the two most important ionospheric characteristics: critical frequency of the F2 layer, *foF2*, and propagation factor for a distance of 3,000 km, *M(3000)F2*.

The 1960s and 1970s were characterized by two different key events for the development of ionospheric science: the use of satellites to explore the circumterrestrial environment and obviously the use of the same for long distance communications. Not of minor importance during the period was the Cold War, which encouraged study of high-latitudes and the exosphere. Chirp sounder systems, capable of real-time identification of the frequency spectrum usable between two points made it clear that the scenario was evolving towards a need for short-term forecasting or an instantaneous picture of the ionosphere, and then real-time frequency management. Short-term forecasting and even nowcasting on a continental scale were among the main objectives of international projects like the European ionospheric COST (Cooperation in Science and Technology) actions, DIAS (DIgital Upper Atmosphere Server) in Europe, and working programs at the institutions including the IPS (Ionospheric Prediction Service) in Australia and the NOAA (National Oceanic and Atmospheric Administration) in the USA, amongst many others.

Real-time management of ionospheric measurements for space weather purposes requires interfacing of hardware systems with the web, which might be relatively simple for GNSS ground-based receiver data, but is much more complicated for ionograms, the output from the large ionospheric vertical sounding network. In addition to the evolution of ionosondes toward more modern digital systems, it was also necessary to develop automatic interpretation or scaling of the ionograms using computer software routines capable of recognizing the images.

While in the past it was only necessary to understand and model the lower part of the ionosphere for HF communications, now an analysis of the entire ionospheric plasma from low levels to the altitudes of GNSS satellites is a vital task for space weather observations and applications. It should also be noted that an instantaneous or forecast picture of the ionospheric plasma it is not only a question of the performance of the geophysical models applied, or simply of the sensitivity of the instruments used, but critically involves the experimental infrastructures on a global or at least regional scale.

HF radio communications use the ionosphere as a medium that is practically infinite, in a way that generates no pollution and is extremely economic in comparison with other systems. They are widely used around the world for important purposes including radio frequencies dedicated to emergency services, strategic implementations, and radio services. However, the ionosphere is a medium subject to solar-terrestrial interactions and disturbed conditions can affect the trans-ionospheric transmissions used by navigation systems. Ionospheric variability may also be considered as evidence of important geophysical phenomena like global climate change and seismic electromagnetic signatures.

## 1.3 Layout of the Book

The introduction provides a conceptual background and brief history of the measurements relevant for ionospheric prediction and forecasting, and their evolution in recent years: from long-term HF planning to nowcast and space weather applications. The key ideas of the different chapters are also summarized. One of the most important issues raised in the book is how to define ionospheric prediction and forecasting in a way that offers maximum utility for guidance in practical radio communication work. Prediction and forecasting are vital areas of research relevant to the management and control of natural environmental systems. They rely on the availability of ionospheric measurements and models appropriate for a particular system. The models are based on theoretical, semi-empirical and/or empirical mathematical-physical approaches, and are governed by a set of parameters, the factors or input variables, the values of which define each possible configuration of the system. An understanding of the genesis and main features of the observable ionospheric parameters is therefore necessary, and this is covered in the first four chapters. The following chapters are dedicated to ionospheric propagation and forecasting. The final Chapter summarizes the main issues described in the book and possible lines of development are presented for the subjects discussed.

Chapter 2 describes ionospheric structure. The Earth's ionosphere is a plasma environment the state of which, at any given time and any specific location, is the result of interactions between multiple physical and chemical processes occurring simultaneously and/or sequentially in different solar-terrestrial domains. The various radiative, chemical, and transport processes in the ionosphere-thermosphere system, together with the effects of solar, interplanetary, and magnetospheric processes above and mesospheric processes below generate: (1) a background ionosphere (climatology-created and affected by incident EUV radiation, processes of ion/electron loss due to neutral constituent chemical reactions, ionospheric electrodynamics and ion drag from thermospheric wind); and (2) a disturbed ionosphere (weather-created and affected by the short term variable impact of the Sun's protons, solar wind particles and/or geomagnetic field). As a consequence, all components of the ionosphere: products, losses, and transport change over multiple time scales, ranging from impulsive solar flares or auroral intensifications (approximately minutes) to solar cycles (approximately 11 years). These produce large scale ionospheric climate and variability changes, storms, auroral and equatorial regions, localized and regional effects, like sporadic E, tilts and gradients, and small-scale ionospheric irregularities, which are summarized in this Chapter. Finally, some attention is dedicated to the highly coupled ionosphere-plasmasphere system together with ionospheric anomalies.

Chapter 3 describes the measurements most relevant for ionospheric prediction and forecasting. An up-to-date review of existing and well known ground-based, in-situ and trans-ionospheric measurement techniques is provided (e.g. vertical and oblique ionosondes, coherent and incoherent scatter radars, the Langmuir probe, and so on) together with the ionospheric electron density height profiles derived from these measurements. Particular emphasis is dedicated to the standard vertical incidence ionospheric characteristics from contemporary coordinated sounding measurements within different global and regional networks. This includes preliminary discussion of the basic physical principles of electromagnetic wave propagation in free space, and a simplified description of magneto-ionic theory with the important concepts of the index of refraction, gyrofrequency, and plasma frequency as applied in ionospheric measurements and monitoring.

Chapter 4 describes how the ionospheric characteristics and parameters derived from measurements and discussed in Chapter 2 exhibit significant differences on timescales ranging from the 11 years of a solar cycle, and even longer, to a few seconds explained in terms of solar cycles, seasonal and daily variations mainly at European mid-latitudes. A brief description of the geographical and geomagnetic variations of these ionospheric characteristics and parameters is also included in this Chapter.

Chapter 5 describes some of the most important electron density models for ionospheric prediction and forecasting purposes. Starting from Chapman's simplified theoretical model, it is shown that the ionosphere can be modelled not only as a superposition of simple Chapman layers, but also in a number of different approaches appropriate for resolving the inherent inverse problem of the bottom-side ionosphere, and the much more complex problems of the topside ionosphere. Since a number of reviews have been published on the topic of empirical and theoretical ionospheric models, special prominence is given here to models that are essential for radio propagation studies. This Chapter amplifies on the two currently most important descriptions of the ionosphere: (1) the International Reference Ionosphere (IRI), which empirically models ionospheric density and temperatures for all latitudes and local times; (2) the NeQuick ionosphere electron density model. Updating models with measured ionospheric characteristics and parameters (model-data-assimilation method) is another important issue addressed in this Chapter.

Chapter 6 describes ionospheric prediction for radio propagation purposes. It outlines research programs that make it possible to predict ionospheric electron density profiles on a global, regional, local, and single station scale in order to perform high-resolution ray tracing of HF radio waves. Contemporary prediction schemes for ionospheric characteristics, like the critical frequency of the F2 layer *foF2,* and propagation factor *M(3000)F2* are presented as long-term, instantaneous, and nowcasting ionospheric specifications for selected locations where measurements are available. In order to extend these to all areas of interest, a number of ionospheric mapping techniques to spatially interpolate derived parameters between sites from the sparse network of measurements (or observations) are described in detail. It is shown that the resulting electron density profiles associated with synoptic mapping can generate an ionospheric prediction model for any location. The highlight of the Chapter is ionospheric mapping, with applications in real-time ionospheric propagation, in which an accurate description of the HF sky wave propagation path and state of ionospheric electron density are required in real-time (e.g. HF sky wave radar).

Chapter 7 describes the electron content observations along a radio signal path between a satellite and a ground receiver station. The sustained growth of Global Navigation Satellite System (GNSS) technologies in numerous implementations means that the effects of the Earth's ionosphere on these satellite-transmitted signals must be studied to achieve the required accuracy for applications like satellite navigation, orbit determination, and satellite altimetry. Throughout this Chapter, regarding total electron content mapping and modelling, important topics related to GNSS signal propagation in the ionosphere, and development of models and warning systems for GNSS applications are addressed. It is shown how new open signals from GNSS, including the modernized US GPS (Global Positioning System), Russian GLONASS (Global Navigation Satellite System), and the planned European Galileo and Chinese COMPASS (Compass Navigation System) systems, can be used to reveal the evolution of space weather events, providing unprecedented opportunities for novel geophysical remote sensing approaches.

Chapter 8 describes ionospheric forecasting. Space weather in the ionosphere has significant impact on ground and space-based communications, the magnitude of which depends on conditions resulting from: (1) variability of solar radiation entering the upper atmosphere (primarily extreme ultraviolet (EUV)); (2) solar plasma entering the Earth's magnetic field; (3) gravitational atmospheric tides produced by the Sun and the Moon; and (4) vertical swelling of the atmosphere due to daytime heating by the Sun. This becomes more intense during solar flare activity, which produces: sudden ionospheric disturbances (SID), ionospheric storms, polar-cap absorption (PCA), and ionospheric storms. Since the resulting ionospheric weather could significantly affect radio wave propagation, timely alert warnings are essential. Therefore, some well-know numerical algorithms are explained for short-term ionospheric forecasting based on statistical and neural network methods. This Chapter also includes a description of forecasting maps, with illustrative examples for comparison of measurements against forecasts.

Chapter 9 contains an overview of the current state of propagation prediction and nowcasting for HF applications and radio links via the ionosphere, which is still important for military, security (government law enforcement agencies), and civilian long-distance communications. This Chapter is an introductory ionospheric propagation "user manual" providing a short description of the geometrical aspects, the basic principles as well as the definition and calculation of the important radio propagation parameters *MUF* (Maximum Usable Frequency) and *LUF* (Lowest Usable Frequency). Existing point-to-point radio links and area HF prediction and nowcasting procedures operated by different national HF radio planning services are briefly presented as examples. Particular emphasis was given to the performance of prediction and real-time frequency management achieved within European projects like the COST actions (238, 251, 271, and 296) related to ionospheric propagation, the European eContent project DIAS, and the European Space Agency (ESA) Pilot Project on Space Weather Applications GIFINT (Geomagnetic Indices Forecasting and Ionospheric Nowcasting Tools), among many others.

Chapter 10 discusses current and future trends in ionospheric prediction and forecasting. In this context the current relevance of ionospheric radio wave propagation is considered in the framework of space weather, as well as the importance of mitigation of solar-terrestrial disturbances and associated signal errors in GNSS and other systems. The recent increased interest in lithosphere-ionosphere coupling is also mentioned because of its effects on human life. The final conclusions are drawn bearing in mind that this book is addressed primarily to those in scientific and engineering communities involved in radio wave propagation in ionized media, radio frequency planning, spectrum management, and system design. New challenges are identified with concrete proposals for lines of development in order to mitigate ionospheric influences on radio systems as far as possible.

## Suggested Readings

Appleton EV (1928) Some notes on wireless methods of investigating the electrical structure of the upper atmosphere (I). Proc Phys Soc 41:43–59

Belehaki A, Cander LjR, Zolesi B, Bremer J, Juren C, Stanislawska I, Dialetis D, Hatzopoulos M (2006) Monitoring and forecasting the ionosphere over Europe: the DIAS project. Space Weather 4:S12002. doi:10.1029/2006SW000270

Belrose JS (1994) Fessenden and the early history of radio science. Radio Sci Bull 5:94–110

Bibl K (1998) Evolution of the ionosonde. Ann Geofis 41:667–680

Bilitza D (2001) International Reference Ionosphere 2000. Radio Sci 36:261–275

Bourdillon A, Cander Lj R, Zolesi B (2009) COST 296 MIERS: Mitigation of Ionospheric Effects on Radio Systems. Ann Geophys 52:209–220

Bradley PA (1995) PRIME (Prediction and Retrospective Ionospheric Modelling over Europe), COST Action 238 Final Report. Commission of the European Communities, Brussels

Chisham G, Lester M, Milan SE, Freeman MP, Bristow WA, Grocott EA, McWilliams KA, Ruohoniemi JM, Yeoman TK, Dyson PL, Greenwald RA, Kikuchi T, Pinnock M, Rash JPS, Sato N, Sofko GJ, Villain J-P, Walker ADM (2007) A decade of the Super Dual Auroral Radar Network (SuperDARN): scientific achievements, new 8 techniques and future directions. Surv Geophys 28:33–109

Gillmore CS (1976) The history of the term ionosphere. Nature 262:347–348

Hanbaba R (1999) Improved Quality of Service in Ionospheric Telecommunication Systems Planning and Operation, COST 251 Final Report. Space Research Centre Printing Office, Warsaw

Hultqvist B (2011) History of EISCAT–Part 1: On the early history of EISCAT with special reference to the Swedish part of it. Hist Geo Space Sci 2:115–121

Jones WB, Gallett RM (1965) Representation of diurnal and geographical variation of ionospheric data by numerical methods. Telecommun J 32:18–27

Lanzerotti L (2005) Assessment of the national space weather program. Space Weather. doi:10.1029/2005sw000202

Manucci AJ, Iijima BA, Lindquister UJ, Pi X, Sparks L, Wilson BD (1999) GPS and Ionosphere. In: Stone WR (ed) Review of Radio Science 1996–1999. Oxford University Press, Oxford

Radicella SM, Leitinger R (2001) The evolution of the DGR approach to model electron density profiles. Adv Space Res 27(1):35–40

Ratcliffe JA (1974) Scientists' reactions to Marconi's transatlantic radio experiment. Proc. IEE 121:1033–1038

Reinisch BW, Galkin IA, Khmyrov GM, Kozlov AV, Bibl K, Lisysyan IA, Cheney GP, Huang X, Kitrosser DF, Paznukhov VV, Luo Y, Jones W, Stelmash S, Hamel R, Grochmal J (2009) The new digisonde for research and monitoring applications. Radio Sci. doi:10.1029/2008RS004115

Zolesi B, Cander LjR (2004) COST 271 Action-Effects of the upper atmosphere on terrestrial and Earth-space communications: introduction. Ann Geophys 47:915–925

# Chapter 2
# The General Structure of the Ionosphere

## 2.1 Introduction

The ionosphere of any planet is defined as that portion of the atmosphere where free electrons and ions of thermal energy exist under the control of the gravity and magnetic field of the planet. There are a number of definitions for the Earth's ionosphere. The main interest of this book is electromagnetic wave propagation though ionospheric plasma as relevant to telecommunications, and so the IEEE Std 211-1997 definition is adopted. The ionosphere of the Earth is therefore taken to be that portion of the upper atmosphere where ions and electrons of thermal energy are present in quantities sufficient to affect the propagation of radio waves.

Most information about the terrestrial ionosphere originated from ionosonde observations at mid-latitudes, and it is reasonable to use this as a basis for describing the behaviour of the ionospheric regions. The physical processes underway in the mid-latitude ionosphere could be assumed to extend over a wide range of latitudes, although significant differences are found at high and low-latitudes due to the orientation of the geomagnetic field. These will be taken into account when relevant for ionospheric prediction and forecasting.

Chapter 2 describes the formation of the Earth's ionosphere and the principal causes for its ionization due to solar–Earth processes and interactions. The main solar phenomena and the Earth space environment including solar wind, the magnetosphere, and the geomagnetic field are briefly presented, with specific emphasis on the importance of the solar and geomagnetic indices for predicting and forecasting the effects on the terrestrial ionosphere.

Then, before discussing the main ionization processes that induce the formation of the ionosphere, there is a short introduction to the neutral atmosphere, its composition, the important parameter of the scale height and the different regions determined by the temperature profile in relation to altitude.

The general structure of the ionosphere and its regular morphology is then described, together with the main mid-latitude ionospheric irregularities.

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_2,

## 2.2 Main Sources of Ionization of the Earth's Upper Atmosphere

### *2.2.1 Sun and Solar Interactions*

#### 2.2.1.1 The Sun's Interior

The Sun is the closest star to the Earth, at a distance varying from 147,100,000 to 149,157,000 km with an absolute magnitude of 4.8, and is classed as a yellow dwarf in the principal sequence of the Hertzsprung-Russell (HR) diagram. The Sun is the most important source of terrestrial upper atmosphere ionization and is illustrated in Fig. 2.1 in a photograph from NASA.

The other main characteristics of the Sun are listed in Table 2.1.

Like similar stars the Sun is composed of about 90 % hydrogen and about 10 % helium. Its activity, i.e. the production of energy, is due to a continuous transformation by nuclear fusion of hydrogen into helium inside the solar nucleus where the temperature is 15 million degrees K.

These transformations can be schematized into three steps as shown in Fig. 2.2. In the first step, Fig. 2.2a, there are two hydrogen nuclei $H^1$, or only protons, which combine to form a *nucleus* of deuterium $D^2$ (1 proton and 1 neutron) plus one positron $e^+$ and a neutrino $\upsilon$, according the following equation:

$$H^1 + H^1 = D^2 + e^+ + \upsilon \tag{2.2.1}$$

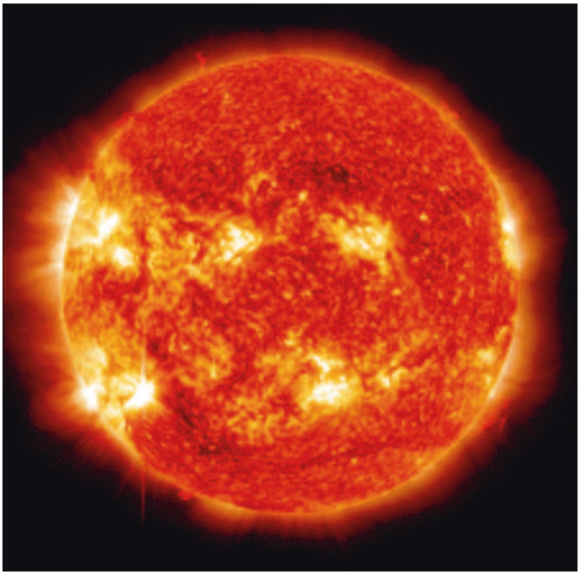


**Fig. 2.1** The image of the Sun captured by NASA's solar dynamics observatory (SDO) with a solar flare on 12 November 2012 peaking at 9:04 p.m. EST. (*Note* that EST is 5 h behind coordinated universal time (UTC). (Credit: NASA/SDO at www.nasa.gov/mission_pages/sunearth/news)

**Table 2.1** Main characteristics of the Sun

| | |
|---|---|
| Solar radius | 695,990 km |
| Solar mass | $1.989 \times 10^{30}$ kg |
| Solar luminosity (energy output of the Sun) | $3.846 \times 10^{33}$ erg/s |
| Surface temperature | 5770 K |
| Surface density | $2.07 \times 10^{-16}$ kg/m$^3$ |
| Surface composition | 70 % H, 28 % He, 2 % (C, N, O, etc.) |
| Central temperature | 15,600,000 K |
| Central density | $150 \times 10^{-9}$ kg/m$^3$ |
| Solar age | $4.57 \times 10^9$ years |

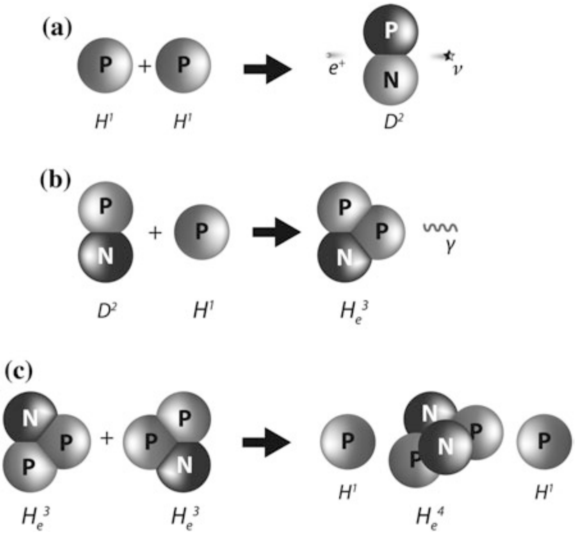


**Fig. 2.2** **a** First step: two hydrogen *nuclei* $H^1$ combine to form a nucleus of deuterium $D^2$ plus one positron $e^+$ and a neutrino $\upsilon$. **b** Second step: an atom of deuterium $D^2$ combines with a proton $H^1$ to form a helium *nucleus* $H_e^3$ plus electromagnetic radiation $\gamma$. **c** Third step: the fusion of two $H_e^3$ *nuclei* into one $H_e^4$ *nucleus* plus two hydrogen atoms $H^1$ with consequent release of energy

Next, the combination of deuterium $D^2$ with a proton $H^1$ forms a helium *nucleus* $H_e^3$ (2 protons plus 1 neutron) plus electromagnetic radiation $\gamma$ (Fig. 2.2b):

$$D^2 + H^1 = H_e^3 + \gamma \tag{2.2.2}$$

Finally, two reactions as described above produce two $H_e^3$ *nuclei*, which can then fuse into one $H_e^4$ *nucleus* (2 protons plus 2 neutrons) plus two hydrogen *nuclei* $H^1$ (1 proton and 1 proton) with consequent release energy (Fig. 2.2c):

$$H_e^3 + H_e^3 = H_e^4 + H^1 + H^1 \tag{2.2.3}$$

The final result is the transformation of 4 protons and 2 neutrons into one atom of helium $H_e^4$, the difference between the initial mass and the final mass is transformed, according to Einstein's law, into kinetic energy.

The energy is radiated from the *nucleus* of 160,000 km in radius as $\gamma$ rays to a distance of 450,000 km from the centre, losing energy and increasing in wavelength to become first X-rays then UV (ultraviolet) rays, and finally visible light.

This region of the solar interior where the energy produced by the inner layers is continuously absorbed and emitted is known as the radiative zone. Above this there is another zone known as the convective zone, where energy is transmitted by the transport of matter. Figures 2.3 and 2.5 provide schematic diagrams of the principal features of the Sun.

### 2.2.1.2 The Solar Atmosphere

The Sun is a gaseous body so it is not easy to define a surface and the borders of its atmosphere. However, it is this external part of the Sun that plays the most important role in the ionization of the Earth's atmosphere, making it possible to predict the resulting influences on ionospheric radio propagation.

What appears from the Earth to be the surface of the Sun, a yellow disk, is the photosphere. As seen in the photo of Fig. 2.1 and in the schematic diagram of Fig. 2.3, this is a very thin layer of about 500–700 km in thickness, and a very small component in the radius of the Sun. Small convection cells are active in this layer and appear on the surface as granulation, or like grains of rice, with diameters of approximately 700–1,000 km and lifetimes of a few minutes. Similar to bubbles from a boiling liquid, the energy is radiated from the photosphere into space. A typical example of this granulation is shown in the photo in Fig. 2.4.

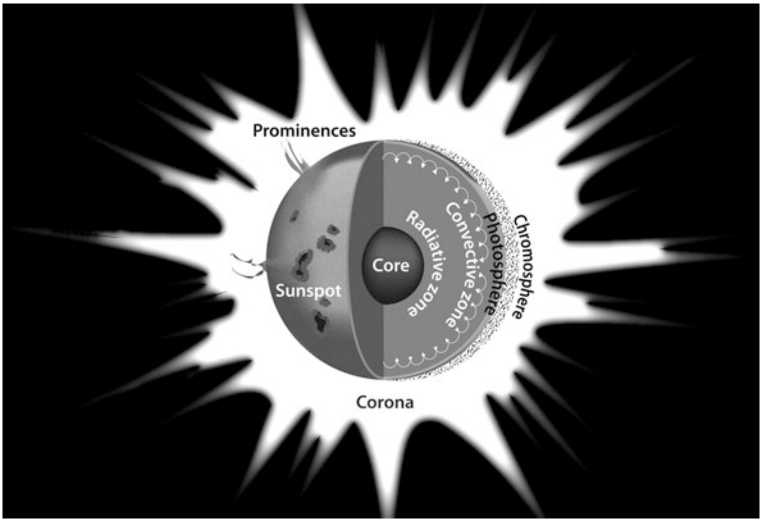


**Fig. 2.3** The Sun: schematic diagram of the internal and external features

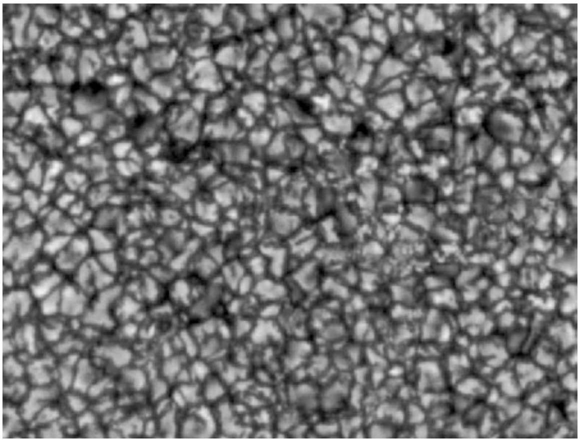


**Fig. 2.4** A magnification of the photosphere reveals the small active convection cells that appear on the surface as granulation or like grains of rice

The effective temperature of the photosphere established by black body physics is about 6,000 K. Most of the energy radiated by the Sun in the range from UV to IR (infrared) rays comes from the photosphere.

### 2.2.1.3 Sunspots and the Solar Cycle

The most important phenomenon occurring in the photosphere, in relation to all other solar phenomena, and for the overall influence and interaction with the terrestrial ionosphere, are certain small regions on the solar surface that appear darker on the solar disk because their temperature is about 3,000 K. These are known as sunspots and are transient phenomena lasting from a few days to some months, of variable dimensions in some rare cases observable with the naked eye, as for example in January 1989. Large sunspots have in fact been observed without optical instruments since antiquity by Chinese astronomers, Greek philosophers like Theophrastus (the pupil and successor of Aristotle also known as the first author of a History of Physics) and on the eve of the modern age in 1610 by Galileo Galilei, who first observed the Sun with a telescope. At the beginning of the 17th century people were surprised to discover that the Sun was not a perfect body but had spots or *maculae*, and observation of the spots and their movement through time proved the differential rotation of the star with a period of about 27 days, faster in the equatorial region at 24 days, and slower in the polar regions at over 30 days. The cause of cooler zones within the sunspot was explained by the existence of strong magnetic fields that obstruct the transfer of hot gases from the lower layers to the photosphere (Fig. 2.5).

The sunspot number has been routinely calculated since 1849 at the Zurich observatory according to the empirical formula defined by its director Rudolf Wolf:

$$R = k(10\,g + s) \tag{2.2.4}$$

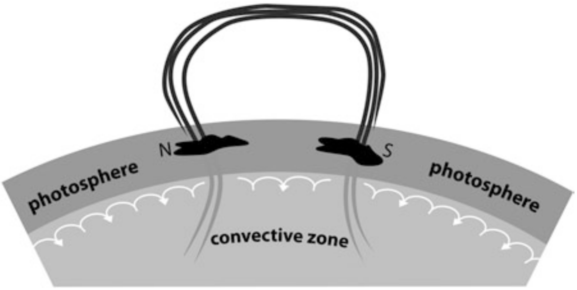


**Fig. 2.5** Strong magnetic fields extend out from the sunspots, which are darker, cooler zones of the photosphere

where $k$ is a correction factor approximately near to 1 taking into account the equipment characteristics, $g$ is the number of groups of sunspot, $s$ is the number of isolated sunspots, and the resulting $R$ is known as the Wolf number. Monthly averages of this number, from 10 to more than 100, and up to 200, reveal a clear 11 year cycle. Although systematic observation only started in the 19th century, it was nevertheless possible to reconstruct all the cycles back to the beginning of the 16th century. Figure 2.6 shows the monthly average of solar sunspot numbers over the last three hundred and 10 years. It is interesting to note that during the period from 1645 to 1715 (the Maunder minimum) as well as at the beginning of the 19th century, the sunspot number remained very low and this behaviour was thought to have some implications for the coincident climate phenomena known as the mini ice-ages. In addition to the main 11 year cycle (not always precisely 11 years but varying from 8 to 14), another 22 year cycle has been observed, taking into account the reversal of the dipolar magnetic field.

The prediction of the behaviour of future solar cycles, naturally only as a smoothed trend, is important in order to predict interactions with the Earth's atmosphere and all the phenomena relevant to space meteorology.

Long-term prediction of solar activity, some years or even some solar cycles in advance, is extremely difficult and at present scientific literature proposes different methods that offer less than adequate results. These methods estimate the extent of the next cycle maximum by considering the length of the previous cycle, the level of activity at sunspot minimum, and the range of the previous cycle. Furthermore, geomagnetic precursors are also considered, for example the variation in the Earth's magnetic field at and before sunspot minimum. The best results are obviously obtained when predicting the remaining part of a current solar cycle 3 years after the sunspot minimum, or by using a method that utilizes an average of different techniques.

#### 2.2.1.4 The Chromosphere, the Corona, and Other Solar Phenomena Including Solar Flares and Coronal Mass Ejections

Just above the photosphere there is a region, about 2,000–2,500 km thick, clearly visible during a total solar eclipse by the moon and reddish in colour: the

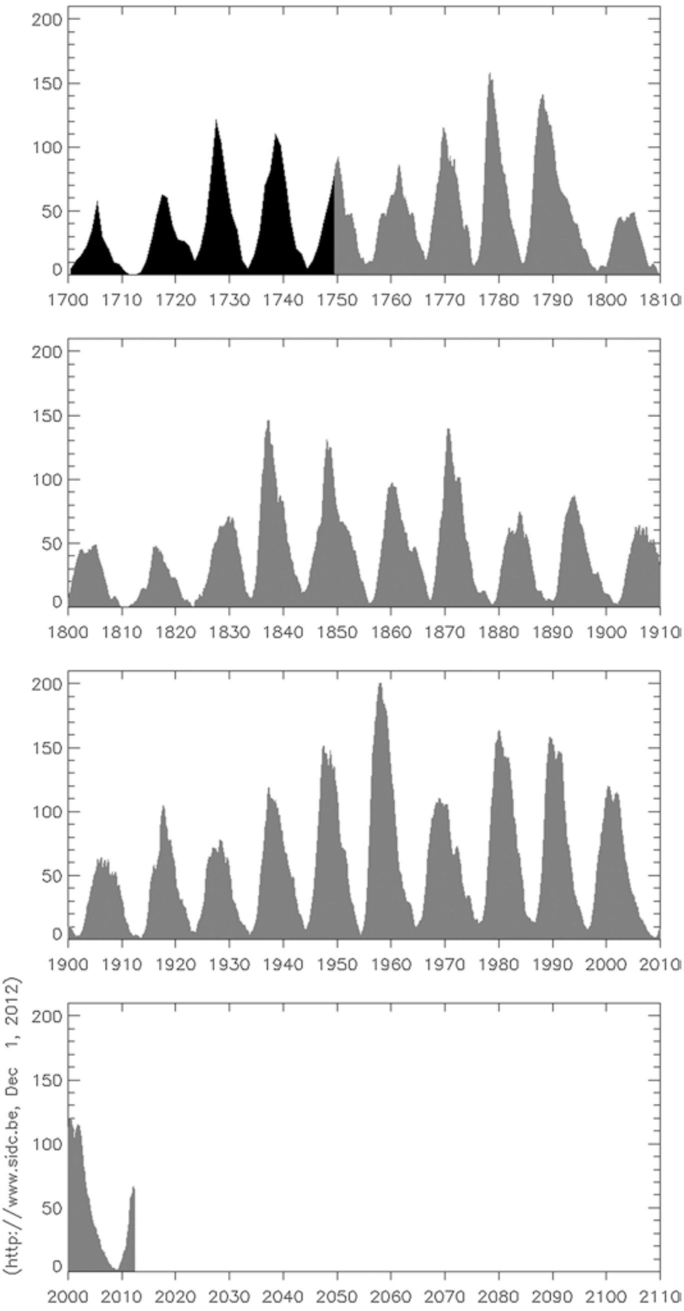


**Fig. 2.6** The yearly and monthly sunspot numbers, from 1700 up to 2012 (http://www.sidc.be)

chromosphere. This region is characterized by progressively increasing temperature from 6,000 to 20,000 K. Within the chromosphere different kinds of plasma eruptions are observed, although in reality they represent different aspects of the same phenomenon. Prominences are emissions of cooler and denser clouds of plasma along the magnetic field lines, visible above a darker zone on the edge of the solar disk known as the limb and lasting for more than one solar rotation. Seen over the solar disk these cooler clouds appear darker and are called Filaments, while *Spiculae* are instead small transient eruptions of material from the surface into the corona region (Fig. 2.3).

Above the chromosphere the corona is known as the outer atmosphere of the Sun, visible during a solar eclipse as white light extending over many solar *radii*. The plasma of the corona is extremely low density but with a temperature of millions of degrees K. Most of the light of the corona comes from the light diffused by the photosphere.

Solar flares are large explosions occurring above the surface of the Sun, close to active regions characterized by the presence of sunspots. They are located along the neutral lines between opposite directions of the magnetic field and produce exceptionally large emissions of electromagnetic $\gamma$ and X-rays, very hot (millions of degrees K) plasma material composed of protons and electrons (Fig. 2.1).

Flares are often associated with Coronal Mass Ejections (CMEs), enormous emissions of plasma in the form of giant bubbles of protons and electrons coming from the corona and flowing along the solar magnetic field. These can last several hours and interact with the solar wind and the Earth's magnetosphere.

All the phenomena occurring in the solar atmosphere result in variations in emissions of electromagnetic energy, and variations in the energy of the particles dispersed into the interplanetary medium. They are directly dependent on the periodic activity of the Sun, as clearly expressed by the presence of sunspots. However, low numbers of sunspots does not necessarily indicate low solar activity or the absence of other solar features.

## *2.2.2 The Solar Wind, the Geomagnetic Field, and the Magnetosphere*

### 2.2.2.1 The Solar Wind

At the end of the 1950s, just before the launch of the first satellites, Sydney Chapman tried to explain the high temperatures measured in the upper atmosphere as a result of heating by hot plasma from the solar corona. The corona was thought to extend throughout the solar system immersing all the planets. As noted in the previous paragraph, the external portion of the Sun's atmosphere, the solar corona, is composed of very hot plasma, from 1 to over 2 million degrees K, a gas with extremely high kinetic energy with the electrons free from the positive *nuclei*.

Starting from this hypothesis Eugene N. Parker proposed a complete explanation for the phenomenon, also taking into account observations of a continuous flux of particles arriving from the Sun. Near the Sun the corona can be approximated as a static atmosphere but with increasing distance from the star the internal pressure of the gas becomes greater than the weight of the upper plasma such as to generate a dynamic flux of particles of increasingly higher velocity. Like a wind these particles flow across the entire solar system, reaching a velocity of hundreds of km/sec. This is called the solar wind and is illustrated in Fig. 2.7.

Together with the continuous flux of particles from the Sun, it is important to note the behaviour of the Solar Magnetic Field, or rather the Interplanetary Magnetic Field (IMF). In the case of a stationary Sun the magnetic field and the solar wind would expand symmetrically along the radius. In reality, solar rotation results in a circular field being superimposed over the radial magnetic field: the result is a spiral form IMF of an intensity decreasing with the square of the distance (see Fig. 2.8). The solar wind is composed of plasma, which being a conductor

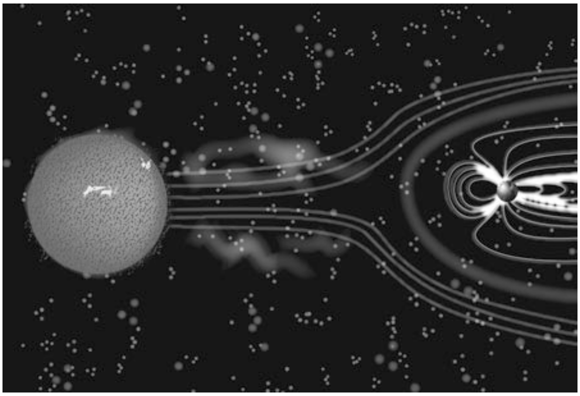


**Fig. 2.7** A pictorial representation of the solar wind

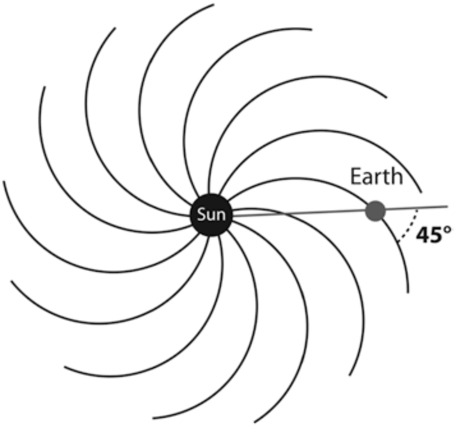


**Fig. 2.8** Spiral form of the IMF, the field lines meet the Earth's orbit at an angle of 45° as confirmed by observations

"freezes" the solar magnetic field, dragging it from the region in which it originated into outer space. This effect is similar to that of flowing water ejected from a garden sprinkler when the individual water drops maintain the radial direction of rotation.

Seen from above on the solar equatorial plane, the IMF appears divided into different magnetic sectors in which the polarity is observed from the Earth as intermittently positive or negative, as shown in Fig. 2.9.

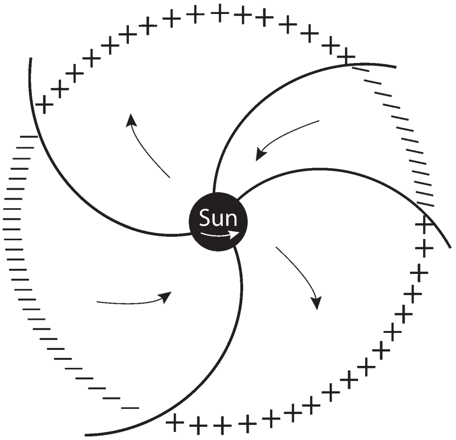
This complex behaviour can be explained by taking into account that the open lines of the solar magnetic field, seen on a plane orthogonal to the equatorial plane, lie parallel and are separated by a neutral current sheet. An observer on the Earth observes opposite polarity over several days when the Earth, along its orbit, is above or below the neutral sheet shaped like a ballerina's tutu (Kelley 1989).

The IMF is a vector of weak intensity varying from 1 to 37 nT near the Earth, with *Bx* and *By* components on the plane of the ecliptic, and *Bz* component perpendicular to it.

#### 2.2.2.2 The Geomagnetic Field

It is important now to schematically describe the geomagnetic field, not only because of its influence over the terrestrial ionosphere, but also to understand how the arrival of the charged particles comprising the solar wind can modify the lines of this field in the region of space around the Earth, generating what is known as the magnetosphere. As an initial approximation the geomagnetic field can be described as the field generated by a simple dipole with its axis slightly offset from the rotational axis. The lines of the field extend outwards from the South Pole and converge inwards towards the North Pole, as was postulated in historical times by William Gilbert in his *De Magnete*, in 1600. This field is the product of different sources which include the dynamics of the metallic fluids within the Earth's core, according to the dynamo theory; residual magnetism in the Earth's

**Fig. 2.9** Different magnetic sectors in which the polarity of the IMF is observed from the Earth as intermittently positive or negative (from Wilcox and Ness 1967)



crust; currents flowing in the ionosphere; and finally the influence of circum terrestrial space through interactions with charged particles in the upper atmosphere. An idealized image of the Earth's magnetic field is shown in Fig. 2.10.

The intensity of the magnetic field is expressed as a vector of magnetic induction **B** and it is measured in Tesla, although it is more practical to use nanoTesla (nT), which are equal to $10^{-9}$ T and also referred to as gamma and equal to $10^{-5}$ gauss, as traditionally used in the CGS system.

The Earth's magnetic field is subject to different categories of variation. There are very slow variations, like the slow secular variation of the magnetic field since the inversion of polarity which occurred around a hundred thousand years ago. There are much faster diurnal variations with a regular trend through the day, and the perturbations induced by solar phenomena or by interaction with the upper atmosphere and magnetosphere.

The vector **F** of the geomagnetic field is described within the reference system as shown in Fig. 2.11 by the so-called seven magnetic elements, 5 of intensity and 2 angular, when $H$ is the horizontal component, $Z$ the vertical component, $I$ the inclination defined as the angle of the vector **F** with the horizontal plane, and $D$ the declination as the angle between the **H** and the geographical meridian:

$$F = \sqrt{X^2 + Y^2 + Z^2} \qquad H = \sqrt{X^2 + Y^2} \tag{2.2.5}$$

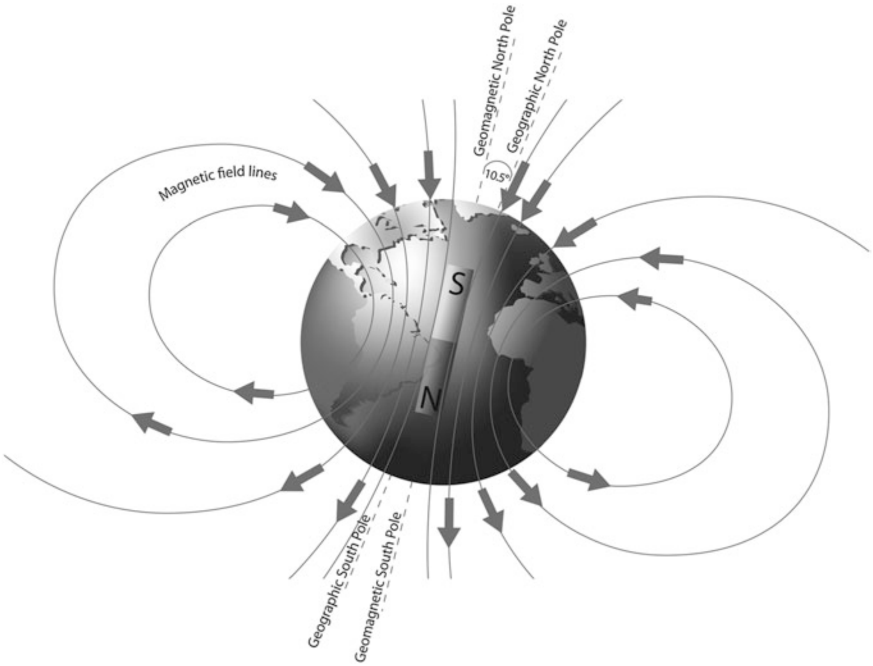


**Fig. 2.10** Idealized view of the Earth's magnetic field as a dipole with a tilt of 10.5° relative to the rotational axis

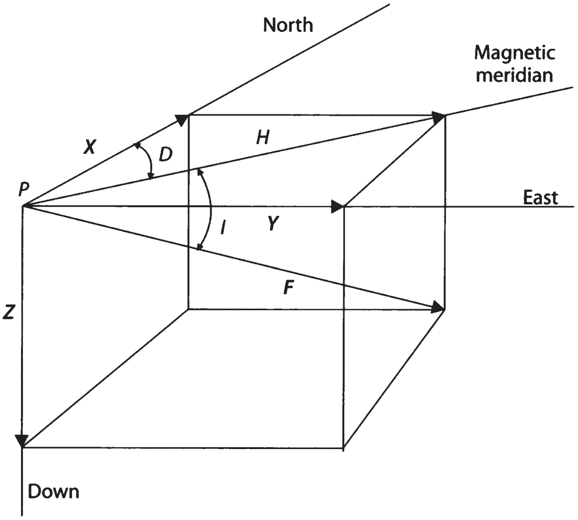


**Fig. 2.11** Elements of the Earth's magnetic field

$$H = F\cos I \quad Z = F\sin I \quad Z = H\tan I \quad X = H\cos D \quad Y = H\sin D \qquad (2.2.6)$$

To analytically represent the geomagnetic field over the entire Earth, it is better to use a reference system with the origin in the centre of the Earth.

A magnetic potential $V$, generated by a dipole at point P with polar coordinates $r$ and $\vartheta$, is:

$$V = \mu_0 \frac{\mathbf{M}\cdot\mathbf{r}}{4\pi r^3} = \mu_0 M \frac{\cos\vartheta}{4\pi r^2} \qquad (2.2.7)$$

when $\mathbf{M}$ is the moment of dipole. The intensity $F$ can be obtained using its gradient or with the equivalent *nabla* operator $\nabla$:

$$\mathbf{F} = -\text{grad } V \quad \text{or} \quad \mathbf{F} = -\nabla V \qquad (2.2.8)$$

See Appendix A for a review of the mathematical operators. The two components along the radius $F_r$ and transversal $F_t$ are obtained by:

$$F_r = -\frac{\delta V}{\delta r} \quad \text{and} \quad F_t = -\frac{\delta V}{r\delta\vartheta} \qquad (2.2.9)$$

$$F_r = 2\mu_0 M \frac{\cos\vartheta}{4\pi r^3} \quad \text{and} \quad F_t = \mu_0 M \frac{\sin\vartheta}{4\pi r^3} \qquad (2.2.10)$$

To analytically represent the geomagnetic field on the Earth's surface it is useful to apply the analytic techniques of spherical harmonics first used by Karl F. Gauss in 1838. The idea is to start from the general physical law also explained by the Maxwell equations, when for magnetic induction **B** the following equivalence applies:

$$\mathbf{B} = -\mathrm{grad}\ V \tag{2.2.11}$$

Considering that in free space, where there are no currents, the Maxwell equation is:

$$\mathrm{div}\ \mathbf{B} = \mathrm{div}(-\mathrm{grad}\ V) = 0 \tag{2.2.12}$$

$$\mathrm{div}\,\mathrm{grad}\ V = \mathrm{div}\ (-\nabla V) = \nabla^2\ V = 0 \tag{2.2.13}$$

where $\nabla^2$ is the *Laplacian* operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.2.14}$$

Applying the *Laplacian* operator to the equation of the geomagnetic potential $V$ in spherical coordinates:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\vartheta}\frac{\partial}{d\vartheta}\left(\sin\vartheta\frac{\partial V}{\partial\vartheta}\right) + \frac{1}{r^2\sin^2\vartheta}\frac{\partial^2 V}{\partial\lambda^2} = 0. \tag{2.2.15}$$

Its general solution may be represented in a simplified form by:

$$V(r, \vartheta, \lambda) = a\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{a}{r}\right)^{n+1} P_n^m(\cos\vartheta)\left[g_n^m\cos(m\lambda) + h_n^m\sin(m\lambda)\right] \tag{2.2.16}$$

where $a$ is the radius of the Earth's surface, $\vartheta$ is the co-latitude, $\lambda$ is the east longitude, $P_n^m$ $(\cos\vartheta)$ are the spherical harmonic Smith functions, and the coefficients $g_n^m$ and $h_n^m$ are the Gauss coefficients. This equation is important because a similar mathematical approach has been used to represent ionospheric characteristics over a sphere in order to obtain global ionospheric maps, as described in Chap. 6. See Fig. 2.12 for an example of a global map of the total magnetic field $F$ in nT for the year 2005.

#### 2.2.2.3 The Magnetosphere and the Plasmasphere

The lines of the geomagnetic field shown in Fig. 2.10 and generated by a simple magnetic dipole can approximate geomagnetic behaviour close to the Earth's surface, but they cannot be extended into space. The exploration of the first satellites during the 1960s revealed a completely different situation at distances of several Earth radii from the planet.

The charged particles of the solar wind flowing from the Sun interact with the lines of the geomagnetic field and change their shape. The lines are compressed on the side facing the Sun and extended on the opposite side in a complex way, generating something resembling the tail of a comet (Fig. 2.13).

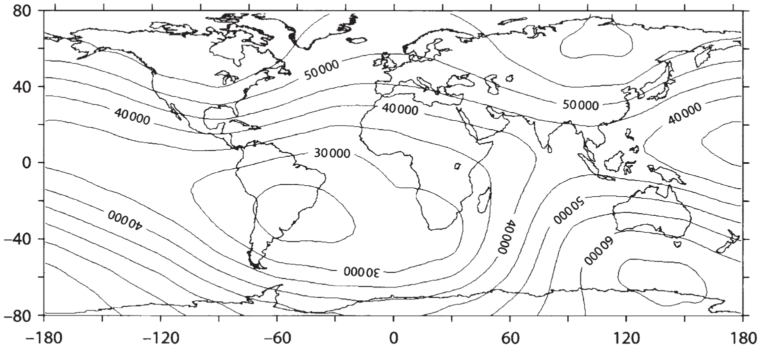


**Fig. 2.12** The Earth's total magnetic field *F* with isolines in nT, year 2005 (from Lanza and Meloni 2006)

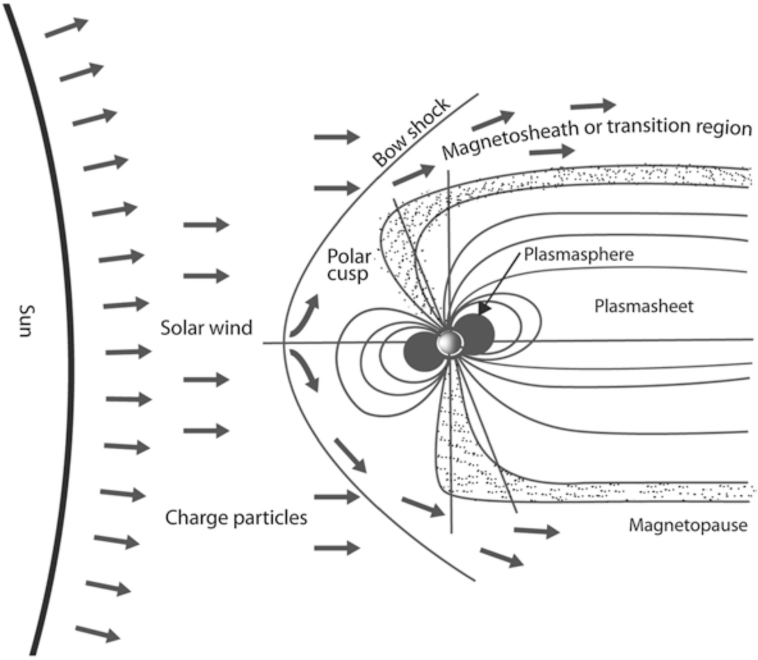


**Fig. 2.13** Diagram of the magnetosphere produced by the interaction of the particles of the solar wind with the geomagnetic field

In the classical and simplified schema of the magnetosphere, the charged particles composing the solar wind start to interact with the lines of the geomagnetic field on the dayside direction producing a shock wave, known-as a bow shock for its shape, which could be considered as the border of the circum terrestrial region. Most of the particles deflected by the bow shock flow along a transition region called the magnetic sheath while a fraction of the particles penetrate into the so-called magnetosphere, the region bounded by the magnetopause. Within this region the charged particles are influenced by the Earth's geomagnetic and electric fields.

There is another region below the lower portion of the magnetosphere called the plasmasphere, which is shaped like a torus around the Earth. The plasmasphere is composed of relatively cool plasma of hydrogen ions and can be considered the upper or outermost portion of the ionosphere, lying above 1,000 km in altitude. This transition region is very influential during measurements of total electron content by satellites, and in all issues affecting Earth-space radio propagation.

The ogive shaped geomagnetic tail extends for over 200 terrestrial radii. The plasma in this portion of the magnetosphere has a different composition: there is an external plasma mantle of very low density and an internal central plasma sheet. The complex behaviour of the particles in the distant regions of the tail are not relevant to the present discussion. It is important to note the two polar cusps at the points where the magnetosphere is in contact with the Earth. Through these two zones there is precipitation of charged particles from the solar wind into the upper atmosphere of the so-called auroral regions.

Finally, the presence of two radiation layers trapped by the fields of the magnetosphere and discovered during space missions is also very important. These are known as the Van Allen Belts. The first layer of radiation is found along the magnetic equator at a distance between 1,000 km to 3 Earth *radii* from the surface of the planet and is composed of protons from the solar wind and from the lower ionosphere. The second layer is at a distance of between 4 and 6 Earth radii and is composed mainly of electrons, again originating from the solar wind. These particles are subject to a rotary movement around the lines of the geomagnetic field as shown in Fig. 2.14, oscillating between North and South. In addition to this oscillation, the electrons and protons are characterized by another movement in opposite directions longitudinally, generating a current called the ring current. On the Earth the effects of this current can be observed and measured as variations in the geomagnetic field.

### 2.2.3 Solar and Geomagnetic Indices

Prediction of solar and geomagnetic activity is essential for forecasting conditions in the near-Earth space environment in general and for ionospheric modelling and propagation processes in particular. This is an important subject in solar-terrestrial studies and significantly depends on the types of solar events and effects described in Sects. 2.2.1 and 2.2.2. For example, solar flares have an immediate and dramatic effect on ionospheric propagation and therefore it is important to provide a useful service for predicting the occurrence of large solar flares. When forecasting geomagnetic activity it is vital to monitor the solar wind upstream of the Earth. It is well established that the critical frequencies of the ionospheric layers depend in a systematic way on measurable quantities related to solar radiation. At the same

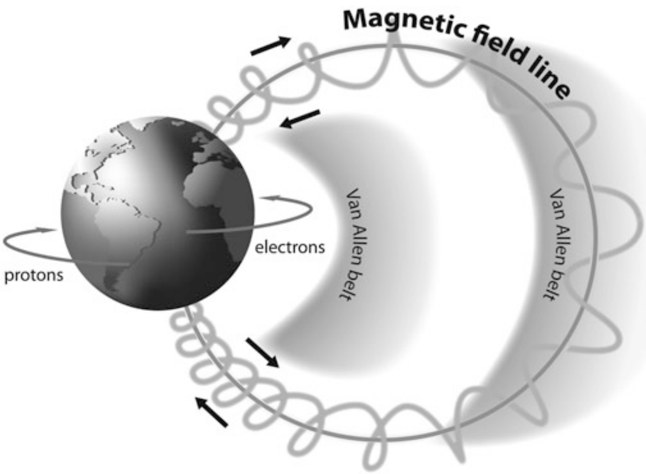


**Fig. 2.14** Electrons and protons in the two layers of radiation known as the Van Allen belts, shaded regions in the figure, oscillate from North to South rotating around the lines of the magnetic field. They simultaneously move longitudinally, generating a current called the annular current

time the structure and dynamics of the ionosphere are profoundly dependent on measurable quantities related to geomagnetic activity.

As described in Sect. 2.2.1, changes in solar activity are cyclic, with one cycle of about 11 years, usually referred to as the 11-year sunspot cycle, another quasi-cycle of about a year, and irregular fluctuations with periods of less than a month. Solar indices such as $R$, $Ri$, $R_{12}$, $\Phi$ and $\Phi_{12}$ are measurable quantities of solar activity or a specific solar radiation. The most widely quoted average sunspot number is the Zurich number ($Rz$) which was replaced after January 1981 with the International Sunspot Number ($Ri$). Figure 2.15 shows the variation of the monthly and smoothed values of the sunspot number $Ri$ for the last five solar cycles including the current 24th cycle.

The 12-month running mean sunspot number $R_{12}$, which considerably reduces complicated rapidly-varying components but does not obscure the slowly varying component, is generally used to study the main component of the solar cycle. The definition of $R_{12}$ is:

$$R_{12}\left(m\right) = \frac{1}{12}\left[\frac{1}{2}R_{m-6} + \sum_{k=m-5}^{k=m+5} R_k + \frac{1}{2}R_{m+6}\right] \tag{2.2.17}$$

where $R_k$ is the mean of the daily sunspot numbers for a single month $k$, and $R_{12}$ is the smoothed index for the month represented by $k = m$. From Eq. 2.2.17 it is clear that the most recent available $R_{12}$ value is centred on a month at least 6 months earlier than the present time. Therefore, $R_{12}$ cannot be used to forecast short-term variations in solar activity.

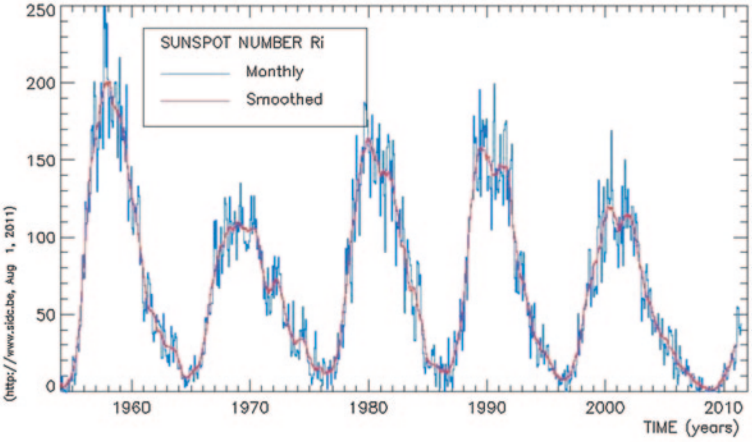


**Fig. 2.15** The monthly and smoothed values of the sunspot number *Ri* from July 1959 to July 2011

Another useful indicator of the level of solar activity is the solar radio noise flux $\Phi$ of 10.7 cm wavelength (frequency 2.8 GHz) which can be regarded as a reference index for dates up to one month. Extended consistent observations of this solar radio noise flux, also known as $\Phi$10.7 flux and expressed in units of $10^{-22}$ Wm$^{-2}$ Hz$^{-1}$, have been conducted by Canadian, Japanese and other laboratories. The relationship between $R_{12}$ and $\Phi_{12}$ was found to be:

$$\Phi_{12} = 63.7 + 0.728\, R_{12} + 8.9\, x\, 10^{-4} R_{12}^2 \tag{2.2.18}$$

The $R_{12}$ and $\Phi_{12}$ indices have generally been adopted by the International Telecommunication Union for long-term and monthly median ionospheric predictions. Measured and predicted values for $R$ and $\Phi$ and their 12-month running mean values ($R_{12}$ and $\Phi_{12}$) are published by different organizations. An example of the results from the Sunspot Index Data Centre (SIDC) in Brussels (http://www.sidc.be/) is shown in Table 2.2.

Geomagnetic indices have been used for a number of years as a suitable technique to represent the complex set of processes associated with the geomagnetic activity discussed in Sect. 2.2.2. Several indices have been applied in solar-terrestrial physics to describe geomagnetic field fluctuations. The most widely-used index is the planetary index 3 h *Kp*. It is derived from the *K* index, which is obtained by observing the *H*, *D*, and *Z* components for eight daily intervals of 3 h (00–03, etc.) at selected geomagnetic observatories. The planetary 3 h *Kp* index is usually expressed in one-third units by adding the signs –, 0, $^+$ to the numbers 0–9, providing a 28-step scale from $0_0$ to $9_0$. The daily $\Sigma Kp$ is the sum of the eight 3 h indices for a universal day. The *Kp* index is extremely useful for evaluating geomagnetically disturbed days (D-days) and the quietest days (Q-days). The selection

**Table 2.2** SIDC news of daily and mean sunspot numbers for April, May, and June 2012 (see http://www.sidc.be/)

SIDC Definitive international and hemispheric sunspot numbers for 2012

| | April | | | May | | | June | | |
|---|---|---|---|---|---|---|---|---|---|
| Date | Ri | Rn | Rs | Ri | Rn | Rs | Ri | Rn | Rs |
| 1 | 42 | 19 | 23 | 69 | 10 | 59 | 85 | 30 | 55 |
| 2 | 51 | 24 | 27 | 68 | 11 | 57 | 106 | 49 | 57 |
| 3 | 46 | 26 | 20 | 69 | 15 | 54 | 116 | 61 | 55 |
| 4 | 41 | 27 | 14 | 62 | 15 | 47 | 104 | 68 | 36 |
| 5 | 34 | 34 | 0 | 63 | 25 | 38 | 106 | 72 | 34 |
| 6 | 34 | 34 | 0 | 62 | 33 | 29 | 107 | 67 | 40 |
| 7 | 19 | 19 | 0 | 57 | 35 | 22 | 89 | 62 | 27 |
| 8 | 16 | 16 | 0 | 61 | 39 | 22 | 75 | 51 | 24 |
| 9 | 9 | 9 | 0 | 65 | 40 | 25 | 74 | 42 | 32 |
| 10 | 10 | 10 | 0 | 69 | 41 | 28 | 85 | 42 | 43 |
| 11 | 19 | 11 | 8 | 79 | 51 | 28 | 90 | 34 | 56 |
| 12 | 37 | 16 | 21 | 72 | 51 | 21 | 78 | 16 | 62 |
| 13 | 35 | 27 | 8 | 75 | 51 | 24 | 79 | 13 | 66 |
| 14 | 33 | 33 | 0 | 85 | 51 | 34 | 86 | 12 | 74 |
| 15 | 58 | 37 | 21 | 87 | 50 | 37 | 82 | 11 | 71 |
| 16 | 39 | 23 | 16 | 98 | 54 | 44 | 74 | 11 | 63 |
| 17 | 51 | 32 | 19 | 79 | 57 | 22 | 62 | 11 | 51 |
| 18 | 72 | 26 | 46 | 73 | 58 | 15 | 40 | 8 | 32 |
| 19 | 95 | 35 | 60 | 78 | 56 | 22 | 36 | 8 | 28 |
| 20 | 108 | 35 | 73 | 83 | 61 | 22 | 21 | 0 | 21 |
| 21 | 104 | 31 | 73 | 79 | 58 | 21 | 11 | 0 | 11 |
| 22 | 94 | 31 | 63 | 57 | 57 | 0 | 11 | 11 | 0 |
| 23 | 89 | 28 | 61 | 55 | 46 | 9 | 11 | 11 | 0 |
| 24 | 86 | 38 | 48 | 62 | 48 | 14 | 14 | 7 | 7 |
| 25 | 81 | 41 | 40 | 67 | 45 | 22 | 12 | 0 | 12 |
| 26 | 75 | 35 | 40 | 57 | 30 | 27 | 25 | 8 | 17 |
| 27 | 68 | 36 | 32 | 59 | 29 | 30 | 49 | 10 | 39 |
| 28 | 72 | 30 | 42 | 78 | 29 | 49 | 61 | 15 | 46 |
| 29 | 71 | 21 | 50 | 56 | 17 | 39 | 72 | 27 | 45 |
| 30 | 67 | 15 | 52 | 52 | 8 | 44 | 73 | 25 | 48 |
| 31 | | | | 64 | 14 | 50 | | | |
| Mean: | 55.2 | 26.6 | 28.6 | 69.0 | 38.2 | 30.8 | 64.5 | 26.1 | 38.4 |

is based on three criteria related to the *Kp* values for each day and the order numbers are selected as the ten quietest and the five most disturbed days of the month.

Another index often used for monitoring geomagnetic activity is the *ap* index. It is a linear index derived from each 3 h planetary index *Kp* in such a way that at 50° dipole latitude the *ap* index approximately represents the maximum disturbance range of the largest of the three components *H*, *D*, and *Z* in units of 2.0 nT. The *ap* index ranges from 0 to 400. The sum of eight 3 h *ap* values gives the daily geomagnetic planetary *Ap* index for a universal day. *Kp* and

*Ap* values are published regularly by the GFZ Helmholtz Centre, Potsdam at http://www-app3.gfz-potsdam.de/kp_index/index.html.

A particular class of fluctuations in the geomagnetic field, with decreases of several hundred nT in the *H* component, are known as geomagnetic, or simply magnetic, storms. Storm variation *D* can be defined as $D = Dst + DS$, where *DS* is due to auroral electrojet activity and *Dst* is due to the ring current in the magnetosphere. Currently, the storm-time variation index *Dst* is calculated using the *H* component from four observatories: uniformly distributed in longitude and remote from the effects of the auroral and equatorial electrojets: Honolulu, San Juan, Kakioka, and Hermanus. Although the continuous hourly values of the *Dst* index primarily give information on geomagnetic storm variations, they are also useful for characterizing persistent quiet-time geomagnetic variations.

Geomagnetic field fluctuations due to the currents in the auroral-zone ionosphere are represented by the *AE* index. The *AE* index is calculated from the *H* components at geomagnetic observatories situated on auroral or sub-auroral latitudes and uniformly spaced in longitude. The maximum positive (upper) amplitude of the *H* component is called the *AU* index and the maximum negative (lower) amplitude is called the *AL* index. The index $AE = AU - AL$, and so represents the difference between the upper and lower limits of the magnetic fluctuations in $\Delta H$ at any time. *Dst* and *AE* values are published regularly by the Data Analysis Center for Geomagnetism and Space Magnetism Faculty of Science, Kyoto at http://wdc.kugi.kyoto-u.ac.jp.

## 2.3 General Atmosphere

### 2.3.1 General Description of the Atmospheric Regions, Composition, and Temperature

The atmosphere is the gaseous envelope surrounding the planet Earth and comprising a mixture of gases. The major components of the atmosphere are summarized in Table 2.3. With the exception of oxygen, the more abundant components are chemically inactive and therefore maintain the same volumetric concentration up to a height of 100–110 km. Inactive components, like carbon dioxide, play an important role in the energy balance, while active components like ozone, despite its very low concentration, play an important role in the absorption of UV radiation and X-rays.

An important property of the atmosphere is pressure *p*, equal to 1.03 Pascal at sea level, and decreasing together with density $\rho$ at increasing altitude. The pressure *p* at an altitude *h* is defined per unit of surface area as the weight of all the air above:

$$p = \int_0^h g\rho dh \tag{2.3.1}$$

where *g* is the acceleration of gravity.

**Table 2.3** Composition of the atmosphere and concentration of the major components

| | | |
|---|---|---|
| *Inactive elements* | | |
| Nitrogen | $N_2$ | 78.11 % |
| Oxygen | $O_2$ | 20.95 % |
| Argon | Ar | 0.93 % |
| Neon | Ne | $18.18 \times 10^{-4}$ % |
| Helium | He | $5.24 \times 10^{-4}$ % |
| *Active elements* | | |
| Water | $H_2O$ | From 0 to 7 % |
| Carbon dioxide | $CO_2$ | From 0.01 to 0.1 % |
| Ozone | $O_3$ | From 0 to 0.00001 % |

Another important parameter characterizing the neutral atmosphere is the scale height $H$ as will be seen in the equations modelling ionization.

The atmosphere is taken to be a fluid in hydrostatic equilibrium composed of a mixture of gases, in which the law of ideal gases is still valid with acceptable approximation:

$$p = nKT \tag{2.3.2}$$

with $K$ the Boltzmann constant ($1.38 \times 10^{-23}$ J/K), $n$ the number of molecules per unit of volume, and $T$ the absolute temperature.

Then, assuming that atmospheric pressure can vary only with height $h$, the upward force due to the gradient is equal to the weight of the atmosphere above:

$$\frac{dp}{dh} = -g\rho \tag{2.3.3}$$

and assuming that the density $\rho$ is also $= n\, m$, with $m$ the mean molecular mass, gives:

$$\frac{dp}{dh} = -g\,n\,m \tag{2.3.4}$$

Then taking $n$ from Eq. 2.3.2:

$$\frac{dp}{p} = -\frac{gm}{KT}dh \tag{2.3.5}$$

By integrating this equation:

$$p = p_0 \exp\left(-\int_0^h \frac{gm}{KT}dh\right) = p_0 \exp\left(-\int_0^h \frac{dh}{H}\right) \tag{2.3.6}$$

When $H = KT/gm$ is defined as the scale height and $p_0$ is the pressure at a height $h = 0$.

To understand the physical meaning of this parameter some approximations are introduced. It is assumed that up to 100 km, $m$ can be considered constant so $H$ should vary only in proportion to $T$ and in the case also of uniform temperature:

$$p = p_0\, e^{-h/H} \tag{2.3.7}$$

So $H$ results as the difference in altitude corresponding to a variation in pressure $p$ equal to $e^{-1}$ i.e. about 0.37 %, in other words the difference in altitude when the atmosphere has a pressure or a density $\rho$ equal to 1/e of the reference level. At sea level, $H$ is equal to 8 km and remains almost constant, as assumed above, up to an altitude of 100 km. Above 100 km the approximation defined above is no longer valid. In fact, in addition to $T$ varying with altitude, $m$ is also not constant but varies according to the molecular mass of the different elements, each of which is distributed according to its own scale height. This means that components with lower molecular mass, like hydrogen or helium, have a greater scale height and are more numerous at higher altitude. Moreover, the scale height represents a measure of the density and pressure gradient, so a lower scale height represents a greater gradient of pressure or density.

The most important parameter used to describe the vertical structure and regions of the atmosphere is the behaviour of temperature with altitude, because this trend reflects the numerous physical and chemical processes in play. As will be discussed later, electron density behaviour can be used to describe the vertical structure of the ionosphere (Fig. 2.16).

From the Earth's surface up to 10 km above the poles and up to 15 km above the equatorial regions the average temperature decrease is about 6.5 ° per km up to

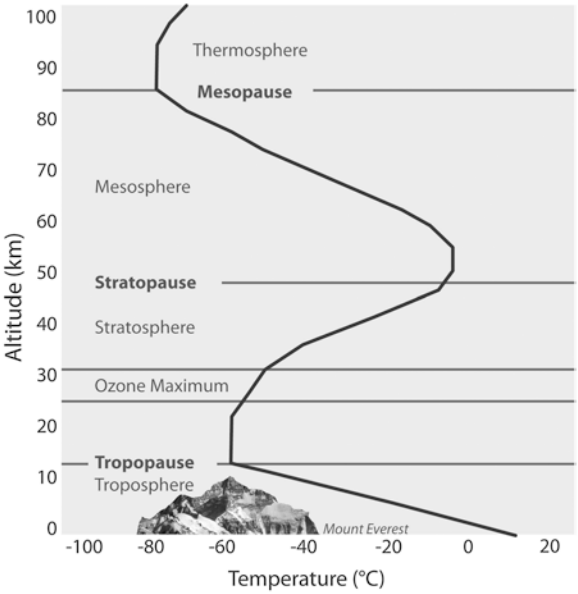


**Fig. 2.16** The temperature of the atmosphere versus altitude and the regions characterized by the inversion of this physical parameter

a level known as the tropopause. Up to this level is the section of the atmosphere that supports life, where most weather phenomena occur, and where the physical process is mainly thermodynamic. The reason for this decrease in temperature is the vertical convective movement of the air: the gas in contact with the warm surface of the Earth expands and cools as it moves upwards in an adiabatic process. Above the troposphere and tropopause the temperature begins to increase again from −70 to 0 K, up to an altitude of 50 km limited by the stratopause. This region is called the stratosphere and the increase in temperature is due to the existence of a layer of ozone, a tri-atomic molecule of oxygen that absorbs UV radiation from the Sun generating new bi-atomic and mono-atomic molecules of oxygen in the process illustrated in Fig. 2.17.

Unlike the troposphere, the temperature inversion blocks vertical movements. Above a maximum altitude of about 50 km the temperature begins to decrease again in a transition region called the mesosphere, continuing up to about 90 km where the temperature reaches about −70 to −90 K. The upper limit of the mesosphere is the mesopause, above which lies the thermosphere where the temperature increases again due to the absorption of UV rays by oxygen and nitrogen, reaching a temperature of 1,500 K at an altitude of 200–300 km. Matter at this height is extremely low density and consequently the temperature parameter is meaningless and it is better to consider the mean quadratic velocity of the particles. In this portion of the upper atmosphere the molecular oxygen and nitrogen are not evenly mixed, instead being stratified according to their molecular mass with a particular scale height. In the thermosphere and upper region of the exosphere the strong UV radiation and X-rays produces an intense ionization so these two regions overlap with the ionosphere.

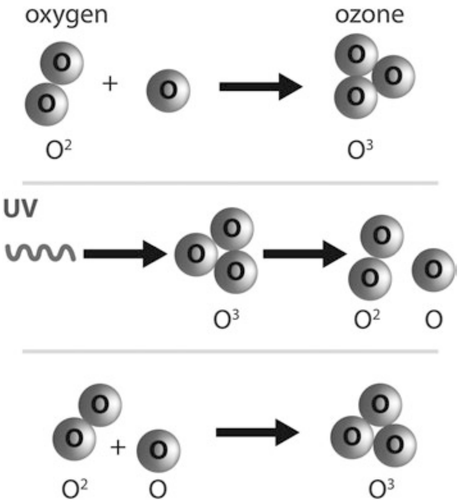


**Fig. 2.17** The stratosphere contains mono-atomic, bi-atomic, and tri-atomic molecules of oxygen. The absorption of UV radiation by an ozone molecule produces an oxygen atom plus a molecule of bi-atomic oxygen. Subsequently the oxygen atom is recaptured by a bi-atomic molecule of oxygen reforming an ozone molecule

## *2.3.2 Formation of the Earth's Ionosphere*

Above the stratosphere, which means beyond an altitude of 50 km, the atmosphere is characterized by a high density of free electrons and free ions, mostly produced by the energetic photo-ionization of UV and X-rays arriving from the Sun, and to a minor extent over high-latitudes by corpuscular ionization. This section of the atmosphere, including the layers referred to as the mesosphere, thermosphere, and exosphere, is defined as the ionosphere.

A high or relevant density of free electrons and free ions is not a clearly defining characteristic because electrons and ions are present at every altitude in the lower and upper atmosphere. Therefore, a more practical definition, originating from the first application of long distance radio communications, is that part of the atmosphere in which the density of ionization is sufficient to deflect radio waves in the 2–30 MHz range.

However, while it is easy to define the lower limit of the ionosphere opinions differ on how to establish its upper limit. A reasonable compromise would be up to 1,000 km, well above any electron density peak and above the limits where the ionosphere is identified as a transition region with the plasmasphere.

### 2.3.2.1 Photo-Ionization

A simplified photo-ionization process is shown in Fig. 2.18, in which a UV radiation strips electrons from neutral atoms which absorb the energy of a photon provided by the quantity $hc/\lambda$, where $h$ is the Plank constant $6.62 \times 10^{-34}$ Js, $c$ the velocity of light, and $\lambda$ the length of the incident electromagnetic wave, as represented by the equation:

$$a + \frac{hc}{\lambda} \rightarrow a^{+} + e^{-} \tag{2.3.8}$$

Here $a$ represents one generic neutral atom, $a^{+}$ represents a positive ion, and $e^{-}$ the free electron, so that the quantity $hc/\lambda$ must be $> Li$ in J, the energy necessary for ionization and related to the specific atomic component $i$.

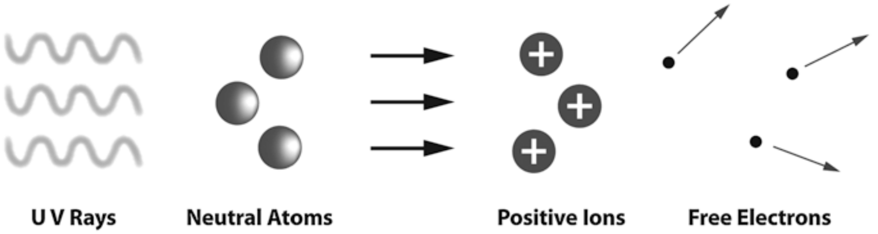


**Fig. 2.18** A simplified scheme of photo-ionization: electrons are detached from their neutral atoms by the energy of UV rays from the Sun, producing ions and free electrons

The amount of energy required for the ionization of the different components of the atmosphere is well known and tabulated. This energy is expressed in electron-volt (eV) or simply with the length $\lambda$ expressed in nm (nanometres) or Angstrom ($\mathring{A} = 10^{-10}$ m) of the associated electromagnetic wave. Furthermore, the production of ion electron pairs is also conditioned by the probability of collision or interaction between a photon and a neutral atom. This is proportional not only to the intensity of radiation $I$, which means the number of photons, but also to the absorption cross section $\sigma_f$ of the ionization which depends on the frequency $f$ for a specific chemical component $i$. The production $q_i$ is given by:

$$q_i = \sigma_f n I \tag{2.3.9}$$

where $n$ is density of the neutral particles of the element $i$.

#### 2.3.2.2 Processes of Electron Loss: Recombination

The reverse phenomenon of photo-ionization is recombination, when the free electrons recombine with the positive ions to produce neutral atoms again. Recombination can occur in two ways. In the first, defined as radiative recombination, the electrons $e^-$ interact directly with the positive ions $a^+$ producing a neutral atom $a$ or molecule plus excess energy $L$:

$$a^+ + e^- \rightarrow a + L \tag{2.3.10}$$

The second and more frequent is called dissociative recombination and involves first an interaction between positive ions $a^+$ originating from the photo-ionization and existing neutral bi atomic molecules $b_2$, like oxygen or nitrogen, according the following equation:

$$a^+ + b_2 \rightarrow ba^+ + b \tag{2.3.11}$$

Substituting in the molecule $b_2$ a neutral atom $b$ with positive ion $a^+$; and then by the combination of free electrons with the positive molecules $ba^+$ producing two neutral atoms:

$$ba^+ + e^- \rightarrow b + a \tag{2.3.12}$$

The probability of occurrence of the first process is proportional to the density of electrons or negative ions $N^-$ and the density of the positive ions $N_i^+$ according to a recombination factor given by:

$$l_r = \gamma_i N^- N_i^+ \tag{2.3.13}$$

where $\gamma_i$ is a coefficient dependent on the specific ion component. This process is peculiar to the higher parts of the ionosphere while, less frequently, in the lower ionosphere the loss of electrons is due to attachment of electrons to neutral atoms, which produces negative ions $a^-$ plus excess energy $L$:

$$e^- + a \rightarrow a^- + L \tag{2.3.14}$$

Again the probability of occurrence is proportional to the electron density $N$ and to the density of the specific element $n$, and also depends on a coefficient factor $\delta_i$ determined by the capability of the element to attract the electrons:

$$l_c = \delta_i N^- n \tag{2.3.15}$$

So finally the equation of continuity or the equation of electronic equilibrium can be written as:

$$\frac{dN}{dt} = q - l + d \tag{2.3.16}$$

where $dN/dt$ represents variation in time and in unit volume of electron density, $q$ is the rate of electron production, $l$ the rate of electron loss, and $d$ a factor that takes into account losses due to diffusion, neutral winds, and vertical electromagnetic drift. Substituting Eqs. 2.3.9, 2.3.13, and 2.3.15 into Eq. 2.3.16, for a specific element $i$:

$$\frac{dN}{dt} = \sigma_f nI - \gamma_i N^- N_i^+ - \delta_i n N^- + d \tag{2.3.17}$$

Then for all the components $i$:

$$\frac{dN}{dt} = \sigma_f nI - \alpha N^- N^+ - \beta n N^- + d \tag{2.3.18}$$

where $\alpha$ and $\beta$ are the effective coefficients of recombination and attachment respectively.

#### 2.3.2.3 The Chapman Model

A simple theory of photo-ionization was formulated by Sydney Chapman in 1931 and is able to describe the behaviour of an ionospheric layer and its variations during the day. The Chapman model was of seminal importance for the many models describing the behaviour of the ionosphere that followed, even in recent times. The mathematical model known as the Chapman function is based on a list of simple assumptions:

1. the atmosphere is composed by only one chemical element exponentially distributed;
2. the atmosphere is plane stratified and not subject to turbulence, diffusion or horizontal variations;
3. the temperature is constant so the scale height $H = KT/gm$ is constant too;
4. ionization is caused only by photo-ionization through absorption of solar monochromatic radiation, and electron loss only by recombination.

According to the above assumptions the Eq. 2.3.18 can be simplified taking into account that $N^- = N^+$ because it is caused by the production of pairs of electron ions, and the contribution of $\beta N^-$ is not considered:

$$\frac{dN}{dt} = \sigma_f nI - \alpha N^2 \tag{2.3.19}$$

Bearing in mind the Eq. 2.3.7 it could be assumed that the number of particles also follows the same law:

$$n = n_0 \, e^{-h/H} \tag{2.3.20}$$

which becomes $n = n_0 \, e^{-z}$ if it is assumed that $z = (h - ho)\,/H$ as a new system of coordinates.

In this system (Fig. 2.19) the decrease in radiation intensity $I$ along the path $ds = \sec\chi dh = \sec\chi Hdz$ is given by:

$$dI = -\varepsilon InH \sec\chi \, dz = -\varepsilon In_0 H \sec\chi \, e^{-z} dz \tag{2.3.21}$$

when $\varepsilon$ is a numerical coefficient of absorption and $\chi$ is the solar zenith angle. It is important to note that the solar zenith angle $\chi$, the angle between the perpendicular to a point on the plane of the Earth's surface and the line of direction of the Sun, can be calculated at any point on the planet for any time of day using the relation:

$$\cos\chi = \sin\varphi \sin\delta + \cos\varphi \cos\delta \cos\omega \tag{2.3.22}$$

when $\varphi$ is the geographic latitude of the point, $\delta$ the declination of the Sun on a given day of the year, and $\omega$ the hourly angle of the Sun. This angle represents true solar time, i.e. the angle between the plane containing the point on the Earth and the Sun and the plane containing the local meridian ($\omega = 0$). Then, integration of the Eq. 2.3.21 between $z = 0$ ando $z = \infty$ gives:

$$I = I_\infty e^{-\varepsilon\, n_0 H \sec\chi \exp(-z)} \tag{2.3.23}$$

where $I_\infty$ may be considered as the intensity of radiation out of the atmosphere or before any absorption. Then considering that $q = \sigma_f nI$:

$$q = \sigma_f n_0 e^{-z} I_\infty e^{-\varepsilon\, n_0 H \sec\chi \exp(-z)} = \sigma_f n_0 I_\infty e^{(-z-\varepsilon\, n_0 H \sec\chi \exp(-z))} \tag{2.3.24}$$

This function shows how $q$ depends on the solar zenith angle $\chi$ and the height $z$. In fact, $q$ is maximum when the solar zenith angle is minimum or $\chi = 0$, i.e. $\sec\chi = 1$:

$$q = \sigma_f n_0 I_\infty e^{(-z-\varepsilon\, n_0 H \exp(-z))} \tag{2.3.25}$$

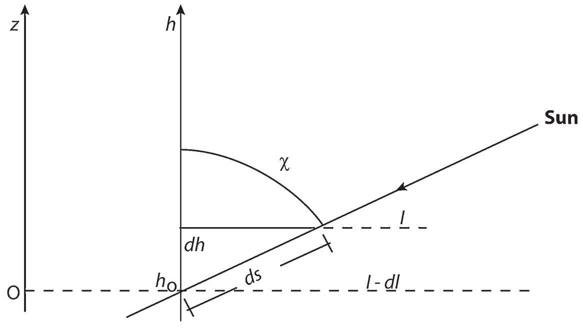


**Fig. 2.19** The decrease in radiation intensity $I$ along the path $dS$ (from Dominici 1971)

Then to establish at which height $z_M$ $q$ reaches maximum it is necessary first to calculate the derivate *dq/dz,* with $\chi = 0$, then imposing the derivate of this function = 0. So:

$$\frac{dq}{dz} = \sigma_f n_0 I_\infty \left[\left(-1 + \varepsilon n_0 H e^{-z}\right) e^{(-z-\varepsilon n_0 H \exp(-z))}\right] = 0 \tag{2.3.26}$$

This is true when:

$$e^{z_M} = \varepsilon n_0 H \tag{2.3.27}$$

If now we consider the height $z = 0$ in the new system of coordinates as the height where $q$ is maximum, true for $\chi = 0$ and now $z_M = 0$, this means that $\varepsilon n_0 H = e^0 = 1$. At this height:

$$q_0 = \sigma_f n_0 I_\infty e^{-1} \tag{2.3.28}$$

so the function in Eq. 2.3.24 could be rewritten for a generic height as

$$q = \sigma_f n_0 I_\infty e^{(-z-\varepsilon n_0 H \sec\chi \exp(-z))} = q_0 e^{(1-z-\sec\chi \exp(-z))} \tag{2.3.29}$$

Then imposing the derivate of this function = 0:

$$\frac{dq}{dz} = q_0 \left(-1 + \sec\chi \cdot e^{-z}\right) e^{(1-z-\sec\chi \exp(-z))} = 0 \tag{2.3.30}$$

which is true if $\left(-1 + \sec\chi \cdot e^{-z}\right) = 0$. This means that:

$$z_M(\chi) = -\ln\cos\chi \tag{2.3.31}$$

Substituting this equation in Eq. 2.3.29 also gives:

$$q_M(\chi) = q_0 \cos\chi \tag{2.3.32}$$

the maximum production rate of ionization $q_M(\chi)$ and its height $z_M(\chi)$ as a function of the solar zenith angle. The consequences of these functions are clearly seen in Fig. 2.20 where it is shown how the quantity $q/q_0$ can vary according to altitude $z$ and varying solar zenith angle $\chi$: the maximum rate of ionization $q_M$ decreases when the Sun goes down to the horizon or moves to its highest position.

### *2.3.3 General Electron Density Profile*

As described in the previous section the principal cause of the high density of free electrons and positively charged ions and molecules is photo-ionization from UV radiation arriving from space and in particular from the Sun, ignoring here other causes like energetic particles, again originating from the Sun, or radioactive elements from natural sources or fall-out pollution. Bearing in mind that the density of neutral atoms decreases with altitude and that ionization becomes weaker because of absorption, the two opposite phenomena should logically result in a layer containing a maximum concentration of free electrons.

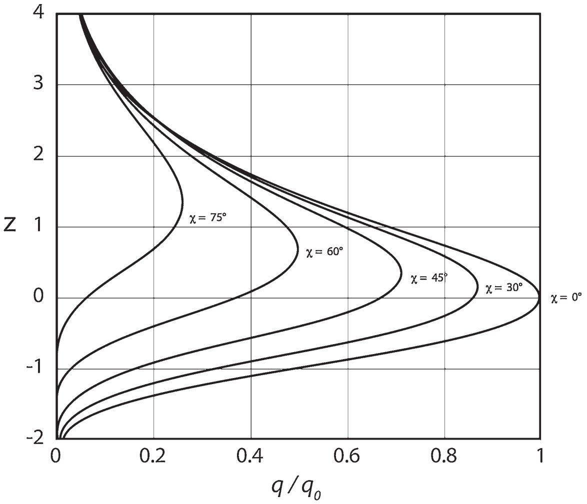


**Fig. 2.20** The quantity $q/q_0$ versus altitude $z$ and for different solar zenith angles $\chi$: the maximum rate of ionization decreases when the Sun goes down to the horizon or moves to its highest position

Since the upper neutral atmosphere is composed of different chemical elements which interact differently with UV radiation, more than one electron density maximum is found. Therefore, in addition to the temperature of the neutral atmosphere, electron density behaviour *versus* altitude is another useful parameter for describing different regions of the ionosphere, with the relative maxima and minima used to identify the ionospheric layers (Fig. 2.21).

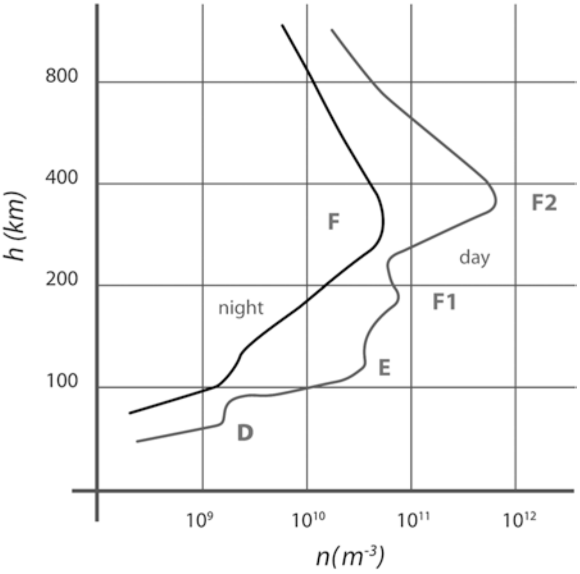


**Fig. 2.21** Diurnal and nocturnal electron density profiles versus altitude

Historically, the letter E was used for the first layer observed by experimental measurements by Edward V. Appleton because of the electric field E reflected by the ionospheric layers, and subsequently the layer above was called F and the layer below D.

The two generic electron density profiles of the figure, one nocturnal and one diurnal, show that during the day all the ionospheric regions can be observed by routine ionospheric soundings and in some seasonal contexts even the splitting of the F region into F1 and F2 layers. Conversely, the nocturnal profile shows the very weak ionizations in the D and E regions. At those altitudes the high density of neutral and ionized particles induces rapid recombination of free electrons and positive molecules, while at higher altitudes of the F region the recombination time is sufficiently long to permit photo-ionization to start again at dawn.

The situation in the polar regions is much more complex because the night can last for several months and a very weak ionosphere is still observed due to marked phenomena of diffusion and transport. Nocturnal profiles logically exhibit a lower density than diurnal profiles at all altitudes.

## 2.4 Ionospheric Regions

The structure and dynamics of the Earth's ionosphere are subject both to large spatial and frequent temporal variations, which can be periodic as well as irregular. The changes that occur in the ionosphere are different at different altitudes because of varying relative ionization, loss, and transport phenomena as described in Sect. 2.3. Consequently, the terrestrial ionosphere is divided into regular and sporadic regions according to the thermal and chemical properties of the neutral gas and ionized components.

### *2.4.1 Regular Ionospheric Regions*

Modern experimental and theoretical investigations divide the ionosphere into three regions: D, E, and F. Under certain solar-terrestrial conditions these regions are split into four main layers D, E, F1, and F2 as shown in Figs. 2.21 and 2.22. In principle, the lower ionosphere (up to 100 km) is a zone in which photochemical processes are the main influence on its formation and ionization balance. The boundary of the middle ionosphere (100–170 km) marks the limit of ionization-recombination processes together with thermal and dynamic processes. The upper ionosphere is the F region, characterized by the transfer of charged particles in plasma by ambipolar diffusion, thermospheric winds, and ionosphere-magnetosphere interactions. The real heights of the ionospheric layers vary with solar zenith angle time, time of day, seasons, solar cycles, and solar activity.

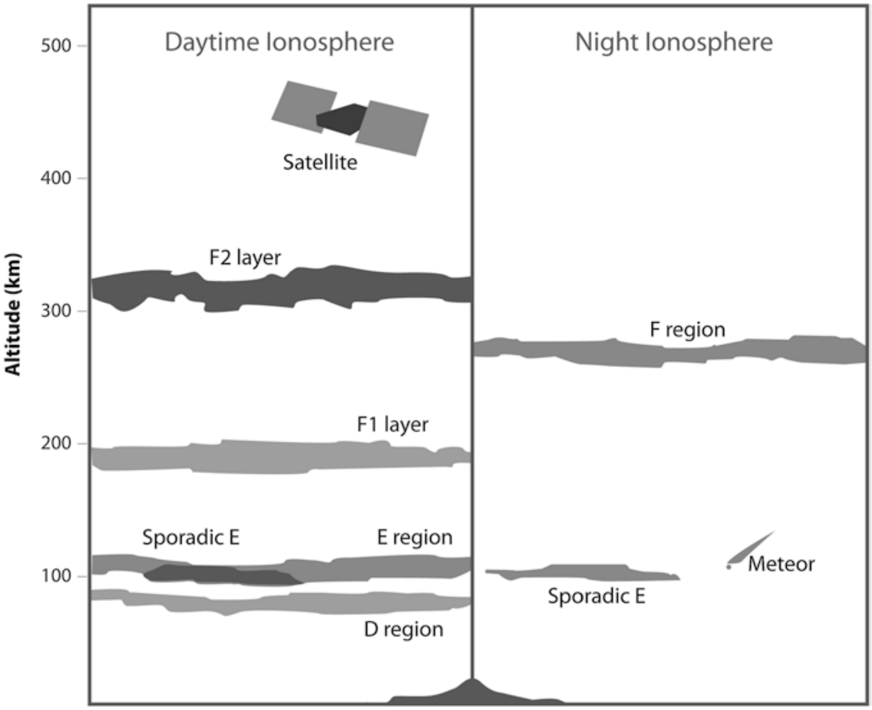


**Fig. 2.22** Day and night structure of the terrestrial ionosphere

#### 2.4.1.1 D Region

The D region of the ionosphere lies at an approximate altitude of 50–90 km, with the section from 50 to 70 km sometimes referred to as the C region and produced mainly by cosmic rays. The neutral component of the D region consists mostly of $N_2$, $O_2$, Ar, $CO_2$, He, and a highly variable quantity of $O_3$ and $H_2O$. The ionized component of the D region consists mostly of $NO^+$ (nitric oxide) as the major positive charge carrier, while electrons, $O_2^-$ and possibly other negative ions are the negative charge carriers (Fig. 2.23). Although the ionizing agents of the D layer are still the object of intense study, solar L-alpha (121.6 nm) radiation appears to be the most important for ionizing NO, solar X-rays (<0.8 nm) for ionizing $N_2$, $O_2$, and Ar, solar ultraviolet (UV) radiation ($\lambda < 111.8$ nm) for ionizing unstable $O_2$, and galactic cosmic rays for ionizing all atmospheric constituents (Fig. 2.24). Rocket and satellite observations suggest that solar cosmic rays, solar protons of 1–100 MeV, and possibly solar electrons of energy >10 keV are the likely contributors to the ionization of the D region.

As Fig. 2.23 shows, D region electron densities are typically around $10^8$–$10^9$ el/$m^3$ depending on height. They are subject to typical diurnal, seasonal, and solar-cycle variations. Since the electron collision frequency in the D region is as high as $2 \times 10^6\,\mathrm{s}^{-1}$, radio wave absorption is quite significant and is very important for the calculation of the Lower Usable Frequency (*LUF*) in radio links as will be discussed in Chap. 9.

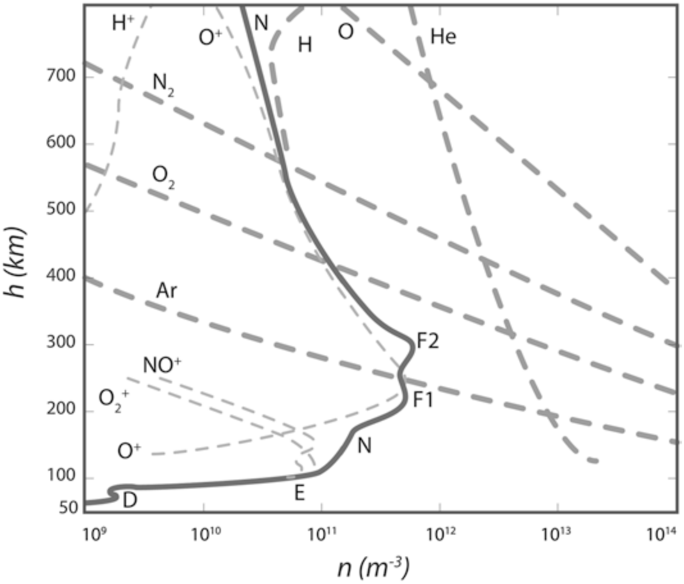


**Fig. 2.23** Principal neutral and ionic constituents of the ionosphere (N is for free electrons) (from Dominici 1998)

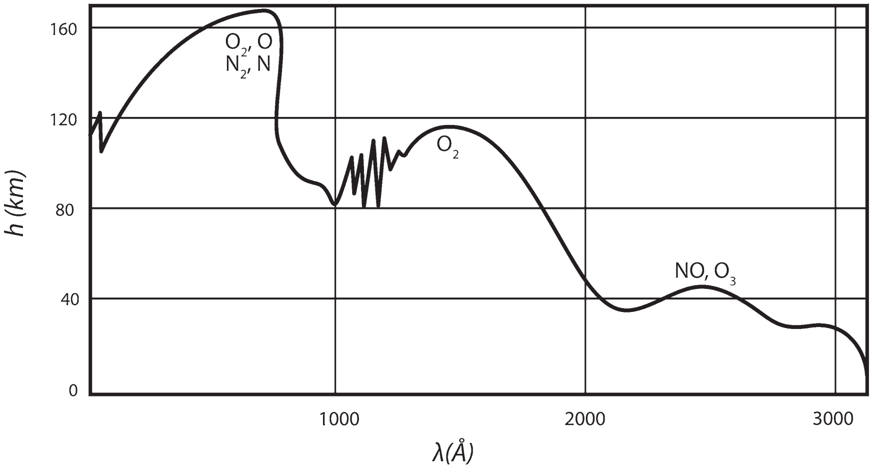


**Fig. 2.24** Atmospheric absorption of solar electromagnetic radiation (from Dominici 1971)

### 2.4.1.2 E Region

The ionosphere from an altitude of 90–140 km is called the E region and contains the E layer, sometimes an E2 layer, and a sporadic Es layer (Fig. 2.22). Since the sporadic Es layer is considered to be an independent phenomenon from the normal E region, it is described in Sect. 2.4.2. Ionization in the E layer is mainly caused by X-rays in the 8–10.4 nm range and UV radiation from 80 nm to L-beta (102.6 nm), producing the principal $NO^+$, $O_2{}^+$ and secondary $O^+$, $N_2{}^+$ ion components (Figs. 2.23 and 2.24). Together with positive ions, the E region also contains electrons with an average E layer electron density of ~$10^{11}$ el/$m^3$ at around ~110 km (Fig. 2.23). The E layer exhibits the Chapman model behaviour with daily maximum at local noon, seasonal maximum in summer, and solar cycle dependence. The E layer does not disappear completely at night and remains weakly ionized. When an E2 layer occurs it is between the normal E and F1 layers at an altitude of about 150 km. Even though the cause of E2 layer formation is still being studied, it seems that the E2 layer is controlled purely by solar ultraviolet radiation.

The electric field and electron drift velocities in the E layer in general, and the low-latitude E layer irregularities arising from the equatorial electrojet in particular, are the key E region features and have been studied in numerous radar experiments.

### 2.4.1.3 F Region

The F region is a major segment of the terrestrial ionosphere and the most important from the point of view of radio communications and navigation systems. It lies between 140 and 600 km in altitude, occasionally with altitudes extending to the upper limits of the ionosphere. As a result of the complex physical mechanisms involved in its formation, solar radiation causes the F region to split into two separate layers, called F1 and F2. The F2 layer is always present during the night and so these layers merge a few hours after sunset, reshaping the F region.

### 2.4.1.4 F1 Layer

The F1 layer of the ionosphere lies at an altitude of approximately 140–210 km (Fig. 2.23). The ionized component of the F1 layer consist mostly of $NO^+$ and $O_2{}^+$ as primary, and $O^+$, $N^+$ as secondary positive charge carriers (Fig. 2.23). The main source of ionization is extreme ultraviolet (EUV) solar radiation in the wavelengths $\lambda \approx 58.4$ and 30.4 nm (Fig. 2.24). The height of F1 varies with solar activity, season, and geomagnetic activity. Rocket experiments have revealed a maximum density height $h_mF1$ from 160 to 180 km, while ionograms from ionosonde measurements indicate $\approx$200 km.

Like the E layer, the F1 layer is a Chapman layer with maximum electron density reaching about $2 \times 10^{11}$ el/m$^3$ at midday (Fig. 2.23). It exhibits dependence on solar zenith angle $\chi$ and sunspot number $R$. Accordingly, the F1 layer is more pronounced in summer than in winter, always disappears during the night and sometimes in winter even during the day. Under those conditions the F1 and F2 layers form a single portion of the ionospheric F region.

#### 2.4.1.5 F2 Layer

The F2 layer is present 24 h a day under all solar-terrestrial conditions, making it the most important layer of the ionosphere (Figs. 2.21, 2.22, and 2.23). The F2 layer has a different nature from the photochemically dominated lower layers and the diffusion dominated topside ionosphere. The main characteristics of the F2 layer are its high variability, on timescales ranging from the 11 years of a solar cycle and even longer, to a few seconds during strong interactions with the plasmasphere above (at altitudes >1,000 km) depending on solar-terrestrial conditions. It is reasonably obvious that replenishment from the plasmasphere plays a significant role in maintaining the F region during the night.

In the F2 layer the dominant charge carriers are $O^+$ ions with secondary $H^+$ and $He^+$ (Fig. 2.23). The main electron–ion source in the F region is the process of photo-ionization by EUV radiation in the wavelength range from 5 to 102.7 nm. The F2 layer has a maximum electron density between $10^{10}$ and $8 \times 10^{12}$ el/m$^{-3}$ (Fig. 2.23). Because the F2 layer is a non Chapman layer, there is no solar zenith angle dependence of the diurnal and latitudinal variations in electron density. However, the Sun does affect the electron density of the F2 layers causing a rapid increase after sunrise, with maximum values occurring at any time during the day.

Moreover, the complexities of F2 layer behaviour can be understood in terms of: (1) chemical changes, because the electron density equilibrium depends strongly on the [O]/[$N_2$] ratio and also on the [O]/[$O_2$] ratio; (2) diurnal heating and cooling, since the height distribution of the $F_2$ layer is governed by the plasma scale height, $H_p = K\,(T_i + T_e)\,/m_i\,g$ (see Eq. 2.3.6); (3) winds in neutral air at heights around 300 km that range between tens and hundreds of ms$^{-1}$, at mid-latitudes directed towards the poles by day and towards the equator at night; and (4) electric fields, since the electrical structure and dynamics of the upper atmosphere generate electrodynamic drifts in the F region that can significantly modify the vertical distribution of ionization. However, because of the dip angle of the magnetic field, the effects of electric fields are much more important for the equatorial ionosphere while the effects of neutral winds are much more important at mid-latitudes.

In the F2 layer, the rate of electron loss depends on the concentration of molecular ions, particularly $O_2$ and $N_2$, whereas the production rate depends on the concentration of atomic oxygen. The photoelectrons produced by ionization processes are hotter than the neutral atoms from which they were formed. This excess energy is gradually shared with the positive ions, but transfer to the neutral component is less efficient. Consequently the temperature of the plasma exceeds that of the neutral air. By day the electrons are considerably hotter than the ions ($T_e > T_i$), but by night the temperatures are more similar. When the plasma is hot it moves up to higher altitudes, where it is effectively stored because of the slower rate of recombination with height. When the plasma cools, at night, it moves down again and so helps to maintain the F region against loss by recombination at lower altitudes.

In 1997 evidence for another layer, named the F3 layer, in the equatorial F region ionosphere was found. Since the presence of such a layer could affect HF communication and surveillance systems, experimental and theoretical investigations were conducted. These suggest that a combined effect of electrodynamic drift and neutral wind during morning to noon periods can drift an F2 layer upwards to form an F3 layer, while the usual photochemical and dynamic effects maintain a normal F2 layer at lower altitudes.

### 2.4.2 Sporadic Ionospheric Layer Es

Although a sporadic E layer occurs at altitudes from 90 to 140 km (the E region), it is usually considered independent of the normal E layer of the ionosphere (Fig. 2.22). The main reason for this view is that Es has very diverse spatial (it can be spread over a wide area or confined to a small area), diurnal (it can appear at any time during the day or night), and seasonal patterns at high, mid, and low-latitudes generated by different physical mechanisms. An Es in low and mid-latitudes occurs mostly during the day and prevalently during the summer months. At high-latitudes, an Es is more likely to occur at night and is frequently associated with aurora. Most importantly, sporadic E layers can have an electron density similar to the F region. However, its random time of occurrence and presence at any particular place makes Es layer prediction very difficult.

There have been many theories for the sporadic E layer based on a large quantity of observational data but the most widely accepted theory still appears to be the wind-shear theory. According to this theory a reversal of the east–west neutral wind combines with the Earth's magnetic field to produce upward and downward motions of heavy ions, thereby compressing them into a layer coincident with the height of wind reversal. Some more recent work indicates that the electric field may play a part in auroral regions where it is very large and the magnetic field is nearly vertical.

## 2.5 Ionospheric Irregularities

Ionospheric irregularities generally include conditions that cannot be accurately described by standard ionospheric models, and events that do not follow conventional patterns predicted on the basis of their physical causes. As illustrated in

Sect. 2.4.1, the F region is highly variable and exhibits, among other irregularities, two very well-known phenomena recognized as travelling disturbances and spread-F.

## *2.5.1 Travelling Ionospheric Disturbances*

Travelling ionospheric disturbances (TIDs) play an important role in the dynamics of the thermosphere and ionosphere. For example, the existence of atmospheric gravity waves (AGWs) in the Earth's upper atmosphere, originating from the weather system of the stratosphere and troposphere, was discovered by Hines (1960) from ground-based observations of travelling ionospheric disturbances. TIDs are irregularities of the F region expressed as wave-like oscillations of the contours of constant electron density, descending slowly through time (Fig. 2.25). TID occurrence and direction of travel show significant diurnal, seasonal, and sunspot cycle variations. They are classified into two types: large-scale TIDs and medium-scale TIDs.

Large-scale travelling ionospheric disturbances (LS TIDs) have a period of 1–3 h and a horizontal wavelength of 1,000–4,000 km, moving faster than 300 m/s. They are thought to be a manifestation of atmospheric gravity waves excited by sources in polar regions of the northern and southern hemispheres. Consequently, the study of LS TIDs could provide important information on auroral processes under

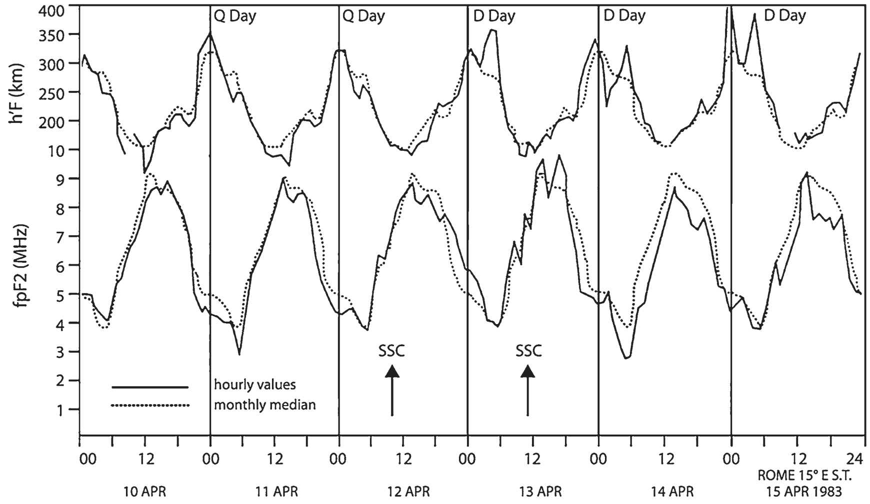


**Fig. 2.25** Variations in some *fpF2* and *h'F* values showing local appearance of TIDs during geomagnetically quiet and disturbed days in April 1983 at the Rome (41.9° N, 12.5° E) ionosonde station

quiet and disturbed geomagnetic conditions. It is known that surges in the auroral electrojet generate trains of large-scale TIDs in both hemispheres, and these trains subsequently propagate towards the equator, causing substantial increases in the height of the ionosphere.

Medium-scale travelling ionospheric disturbances (MS TIDs) have shorter periods of 10 min to 1 h and a wavelength of up to 300 km, moving more slowly (50–300 m/s). MS TIDs are related to meteorological phenomena like neutral winds or the solar terminator, which produce AGWs expressed as TIDs at ionospheric heights.

### 2.5.2 Spread-F

Usually the F region reflections of radio waves are clear echoes indicating well-defined ionospheric layers as described in Sect. 2.4. Spread-F occurs when the ionospheric F region becomes diffused and irregularities scatter the radio waves (Fig. 2.26). The size of F region irregularities, which occur in patches, varies from 20 km to more than 100 km. The solar cycle and seasonal variations, as well as local time dependence, are well established characteristic of spread-F occurrence. In addition, the probability of spread-F occurring is strongly dependent on latitude since spread-F occurrence rates and characteristics at low-latitudes differ significantly from high-latitudes.

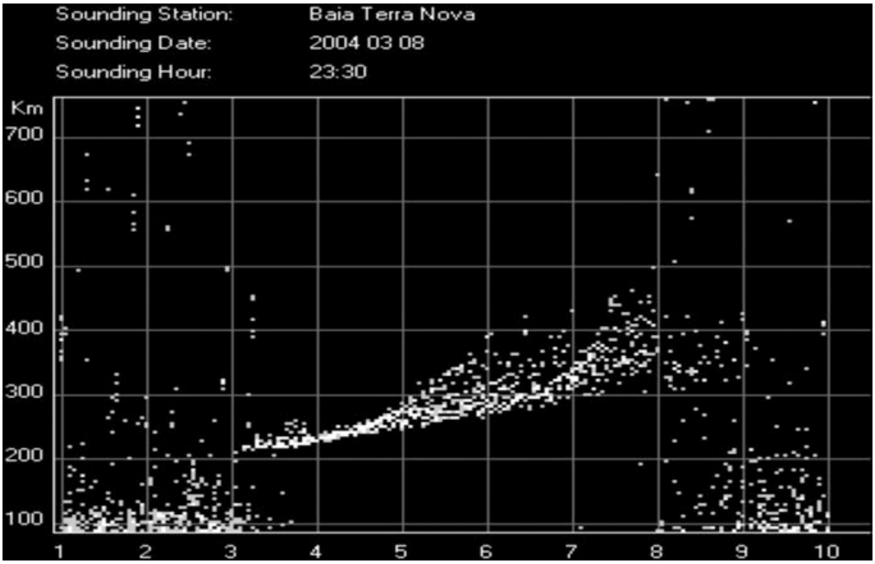


**Fig. 2.26** A typical spread-F ionogram recorded at an Antarctica Baia Terra Nova (74.7° S, 164.1° E) ionosonde station at 2330 UT on 8 March 2004

At low-latitudes, spread-F occurs mostly during the night and around the equinoxes. At mid-latitudes, spread-F is more likely to occur at night and in winter. At latitudes greater than about ±40° geomagnetic latitude, spread-F is a night-time phenomenon, mostly around the equinoxes. Near the magnetic poles, spread-F is frequently observed both during the day and night. At high-latitudes spread-F is correlated with geomagnetic activity. On the contrary, at the equator there is no geomagnetic control over spread-F.

## Suggested Readings

Akasofu S.-I, Chapman S (1972) Solar Terrestrial Physics. Oxford University Press, Oxford

Balan N, Bailey GJ, Abdu MA, Oyama KI, Richards PG, MacDougall J, Batista IS (1997) Equatorial plasma fountain and its effects over three locations: evidence for an additional layer, the F3 layer. J Geophys Res 101:2047–2056

Boschi E, Visconti G, Rizzi V, Zolesi B (1991) Il Buco dell'ozono: indizio del cambiamento globale. Editor: Il Cigno Galileo Galilei

Bowman GG (1964) Spread-F in the ionosphere and the neutral particle density of the upper atmosphere. Nature 201:564–566

Bowman GG (1990) A review of some recent work on mid-latitude spread-F occurrence as detected by ionosondes. J Geomag Geoelectr 42:109–138

Brekke A (1997) Physics of the Upper Atmosphere. Wiley published in association with Praxis Publishing

Campbell WH (1997) Introduction to Geomagnetic Fields. Cambridge University Press, Cambridge

Davies K (1990) Ionospheric Radio. IEE electromagnetic waves series 31, Peter Peregrinus Ltd, London

Dominici P (1971) Radio Propagazione Ionospherica. Supplement to vol XXIV of Ann Geofis

Dominici P (1998) My first fifty years in ionospheric research. Ann Geofis 41:857–883

Dominici P, Cander LjR, Zolesi B (1997) On the origin of medium-period ionospheric waves and their possible modelling: a short review. Ann Geofis XL:1171–1178

Hargreaves JK (1992) The Solar-Terrestrial Environment. Cambridge Atmospheric and Space Science Series 5, Cambridge University Press, Cambridge

Hines CO (1960) Internal gravity waves at ionospheric heights. Can J Phys 38:1441–1481

IEEE Std 211-1997, IEEE standard definition of terms for radio wave propagation. The Institute of Electrical and Electronics Engineers, New York (1998)

ITU-R Choice of indices for long-term ionospheric predictions, Recommendation ITU-R P. 371-8, International Telecommunication Union, Geneva (2001)

Kelley CM (1989) The Earth's Ionosphere. Academic Press Inc, California

Lanza R, Meloni A (2006) The Earth's Magnetism. Springer, Heidelberg

McNamara LF (1991) The Ionosphere: Communications, Surveillance, and Direction Finding. Krieger Publishing Company, Malabar

Sugiura M (1964) Hourly values of equatorial Dst for the IGY. Ann Int Geophys Year 35:9, Pergamon Press, Oxford

Villante U (2001) Al di là delle nuvole, la fisica delle relazioni Sole-Terra. Bollati Boringhieri, Saggi Scienze

Whitehead JD (1989) Recent work on mid-latitude and equatorial sporadic-E. J Atmos Terr Phys 51:401–424

Wilcox J, Ness NF (1967) Solar source of the interplanetary sector structure. Solar Phys 1:437–445

## Additional Web Sites

http://www.astronomynotes.com/
http://solarscience.msfc.nasa.gov/
http://www.windows2universe.org/
http://www.daviddarling.info/encyclopedia/ETEmain.html

# Chapter 3
# Ionospheric Measurements and Characteristics

## 3.1 Introduction

The most important techniques systematically used for monitoring the ionized layers of the Earth's upper atmosphere are based on propagation effects that influence radio waves travelling through the ionosphere (Sect. 3.2). A review of the principles of electromagnetic wave propagation in free space, starting from the basic Maxwell equations, is briefly presented in Sect. 3.2.1. A simplified description of magneto-ionic theory is then presented in Sect. 3.2.2. This is useful for describing propagation in an ionized medium like the ionosphere, with the important concepts of the index of refraction, gyrofrequency, and plasma frequency to be applied in ionospheric measurements and monitoring.

Ionosondes and HF radar systems, acting both vertically and obliquely, have been the most popular instruments for ionospheric measurements over the last eight decades (Sect. 3.3). The principal technical characteristics of the equipment used for vertical and oblique incidence sounding, the most important and widely used methods for routine ionospheric observations, are described in Sect. 3.3 together with the basic physical theory of this method derived from magneto-ionic theory. The same Section also introduces the schematic operation of a standard ionospheric station, its outputs (ionograms) and their interpretation and data presentation.

Although, there are more than 200 ionosondes available worldwide, their spatial and temporal coverage of ionospheric observations is limited. By contrast, today, there are thousands of ground-based receivers of the Global Navigation Satellite System (GNSS) deployed to monitor the Earth's surface, and these have also been used to simultaneously and continuously monitor the ionospheric total electron content (*TEC*). Apart from these ground-based techniques with global coverage, there are also HF (SuperDARN) and VHF (STARE) coherent backscatter radar systems, and incoherent scatter radar systems, with a few facilities available for investigating the ionospheric plasma properties at different sites so that a regional perspective can be obtained with extremely high precision (Sect. 3.4). Since the beginning of space exploration with rockets and satellites, numerous

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_3,

missions have contributed to the knowledge of the ionosphere with in-situ measurements of plasma properties under different solar-terrestrial conditions, and excellent imaging of the ionosphere from space (Sect. 3.5). From the first experiments up to the modern era, the focus has always been on complementary measurements at the same location or multiple instruments in different locations, with strong international collaboration for extensive data collection, dissemination and research defined by the synoptic nature of the phenomenon.

## 3.2 Basic Physical Principles of the Magneto-Ionic Theory

The study of the refraction and reflection of an electromagnetic wave propagating within a neutral gas, when there is also a relevant presence of positive ions and free electrons subject to an external magnetic field, started just after O. Heaviside and A. E. Kennelly proposed their interpretation of the first radio link over the Atlantic Ocean. First W. H. Eccles and J. Larmor considered the problem and then the theoretical aspect was finally expounded in complete form by E. A. Appleton and D. R. Hartree, arriving simultaneously and independently at the identical conclusion, and consequently the equation describing the index of refraction of an electromagnetic wave in the ionosphere is universally known as the Appleton–Hartree equation. Also worthy of note are the contributions of two other scientists to this work, S. Goldenstein and W. Altar.

The aim of this Section is to provide a simplified outline of the basic physical principles of magneto-ionic theory, and how the general Appleton–Hartree equation makes it possible to introduce important concepts like the frequency of plasma, the gyrofrequency, and their influence on the ionospheric measurements described in Sect. 3.3. However, for a complete and exhaustive treatment of the theory see the seminal reference works: “The Magneto-Ionic Theory and its Applications to the Ionosphere” by J. A. Ractliffe, and “Radio Wave in the Ionosphere” by K. G. Budden. Short but rigorous treatments are also provided in the volumes by K. Davies and by P. Dominici.

### 3.2.1 Basic Equations of an Electromagnetic Wave

The logical starting point for an introduction to the Appleton–Hartree equation are the basic Maxwell equations describing the propagation of an electromagnetic wave in space at finite velocity and determining energy transport. These equations apply to free space and an isotropous undisturbed homogeneous medium, and can be written as follow:

$$\operatorname{div} \mathbf{E} = 0 \tag{3.2.1}$$

$$\operatorname{div} \mathbf{B} = 0 \tag{3.2.2}$$

$$\mathrm{curl}\ \mathbf{E} = -\delta\mathbf{B}/\delta t \tag{3.2.3}$$

$$\mathrm{curl}\ \mathbf{B} = \varepsilon_0\ \mu_0\ \delta\mathbf{E}/\delta t \tag{3.2.4}$$

with **E** the vector of the electric field in V/m, **B** the vector of magnetic induction in Tesla, $\varepsilon_0$ the electric permittivity, and $\mu_0$ the magnetic permittivity.

As is well known, these equations were formulated by J. C. Maxwell in order to explain the experimental physical laws enunciated during the 19th century. The first equation, from the Gauss theorem, concerns the fact that the flux of the electric field **E** within a closed surface depends only on the electric charge content and, in the case of a vacuum, this is equal to 0. Similarly, in the second equation, the flux of **B**, the vector magnetic induction, is 0 because the field lines do not come from an isolated source but connect one pole to another. The third equation regards the law of Faraday-Neumann-Lenz concerning the induction of an electric field by variation in time of the vector of the magnetic induction **B**. Finally, in the fourth equation, the Ampere theorem, a variation in time of the vector **E** produces a magnetic field of intensity $B/\mu_0$. Now applying the operator *curl* (see Appendix A) to the Eq. 3.2.3:

$$\mathrm{curl}\,\mathrm{curl}\ \mathbf{E} = -\mathrm{curl}\ \delta\mathbf{B}/\delta t$$

and considering the Eq. 3.2.4:

$$\mathrm{curl}\,\mathrm{curl}\ \mathbf{E}\ = -\varepsilon_0\ \mu_0\ \delta^2\ \mathbf{E}/\delta t^2 \tag{3.2.5}$$

Then taking into account that:

$$\mathrm{curl}\,\mathrm{curl}\ \mathbf{E} = \mathrm{grad}\,\mathrm{div}\ \mathbf{E} - \mathrm{div}\,\mathrm{grad}\ \mathbf{E} \tag{3.2.6}$$

and that for Eq. 3.2.1 grad div $\mathbf{E} = 0$:

$$\mathrm{curlcurl}\ \mathbf{E} = -\nabla^2\mathbf{E} \tag{3.2.7}$$

where $\nabla^2$ is the *Laplacian* and $\nabla$ is *nabla* operator so:

$$\nabla^2\mathbf{E} = \varepsilon_0\ \mu_0\ \delta^2\mathbf{E}/\delta t^2 \tag{3.2.8}$$

By using the same method similarly:

$$\nabla^2\mathbf{B} = \varepsilon_0\ \mu_0\ \delta^2\mathbf{B}/\delta t^2 \tag{3.2.9}$$

Each of these two vectorial equations can be written along the three coordinate axes, so considering that in a homogenous and *isotropus* medium both **E** and **B** can vary only along the direction $x$, choosing a reference system where $B_y = 0$ and $E_z = 0$ they can be rewritten as follows:

$$\nabla^2 E_y = (1/c^2)\ \delta^2 E_y/\delta t^2 \tag{3.2.10}$$

$$\nabla^2 B_z = (1/c^2)\ \delta^2 B_z/\delta t^2 \tag{3.2.11}$$

These equations represent the well known equations for a wave propagating along the $x$ axis with phase velocity $c = 1/(\varepsilon_0\ \mu_0)^{1/2}$ corresponding, for a vacuum, to the velocity of light. The general solution of these equations is for $E_y$:

$$E_y\,(x,t) = F\,(x - vt) - G\,(x + vt) \tag{3.2.12}$$

where $F(x - vt)$ represents a wave going forwards and $G(x + vt)$ represents a wave going backwards while $F$ and $G$ are the functions of the arguments in brackets.

As an example of a specific solution we can take the harmonic wave:

$$E_y = A\sin\,(\omega t - \psi) \tag{3.2.13}$$

where the dependence on time is a sinusoidal function. In this case, the wave is only progressive and so the backwards wave is 0. The operational function $F$ is a *sinusoid*, $A$ is the amplitude, and $(\omega t - \psi)$ is the phase at the instant $t$ and at distance $x$ from the origin, $\omega = 2\pi/T$ the pulsation or the variation of the phase by unit of time, $\psi$ the initial phase at distance $x$, and $t = 0$. This quantity can be written as $\psi = (kx + \varphi)$ where $\varphi$ is the phase at the origin, $x = 0$ and $t = 0$, and $k = 2\pi/\lambda$ is the wave number, which is the change of phase proportional to the distance $x$. In this case, the Eq. 3.2.13 can be written:

$$E_y = A\sin(\omega t - (kx + \varphi)) \tag{3.2.14}$$

or in the complex form (see Appendix A).

Then, if we choose the reference time such that $\varphi = 0$, the Eq. 3.2.14 can be written as:

$$E_y = A\sin(\omega t - kx) \tag{3.2.15}$$

This regards the simple case of a plane wave linearly polarized, meaning that the front of wave is flat and that, with the components $E_z = 0$ and $B_y = 0$, the plane of oscillation of the vector **E** is orthogonal to the plane of oscillation of the vector **B**. The two planes of oscillation are orthogonal to the direction of propagation $x$ (Fig. 3.1).

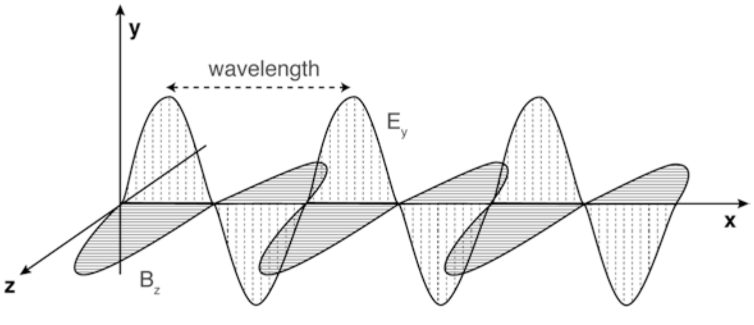


**Fig. 3.1** Schematic view of the propagation of an electromagnetic wave

In general, a linearly polarized plane wave can also be considered the result of two harmonic waves linearly polarized over two planes orthogonal to each other and propagating in the same direction and with the same frequency as follows:

$$E_z = A_z \sin(\omega t - \psi_z) \tag{3.2.16}$$

and:

$$E_y = A_y \sin\left(\omega t - \psi_y\right) \tag{3.2.17}$$

The result is still a harmonic wave linearly polarized if $\psi_z = \psi_y$ or it is an elliptic or circular polarization considering the difference of the phases and the difference of the amplitudes $A_y$ and $A_z$.

Finally, to complete this short review of the basic equations of an electromagnetic wave it is important also to introduce the concept of velocity of phase and velocity of group. Up to this point, an ideal monochromatic wave was considered, and in this case, there is no difference between the velocity of phase and the velocity of group. In nature, there are no simple monochromatic waves, especially if propagating in a medium. According to the Fourier theorem, a given non harmonic wave may be considered the resultant of $n$ monochromatic waves, each with frequency and phase as follows:

$$E_y = a + \sum_{n=1}^{\infty} An \sin(n\omega t - \varphi_n) \tag{3.2.18}$$

where $a$ is the average value of $E_y, A_1 \sin(\omega t - \varphi_1)$ the first or fundamental harmonic, and the others for $n = 2, 3 \ldots$ are the second, third, etc. Taking this analysis into account a generic wave propagating in a medium could be considered as a resultant wave obtained from the composition of single monochromatic waves. In this wave, the pulsation and initial phase are the average of the pulsations and the initial phases of the wave components, while the amplitude will vary with cosine of a different phase. In the particular case of two monochromatic waves, by applying the trigonometric formulas, the phase velocity is:

$$v_p = (\omega_1 + \omega_2)/(k_1 + k_2) \tag{3.2.19}$$

while the group velocity is:

$$v_g = (\omega_1 - \omega_2)/(k_1 - k_2) \tag{3.2.20}$$

The resultant wave of two monochromatic waves is a wave of amplitude oscillating at a velocity given by $v_g$. This is the velocity at which a group of waves, propagating at velocity $v_p$, are enveloped together (Fig. 3.2). Instead, in free space the velocity of propagation of the two waves does not depend on frequency and so they are the same, with $v_g = v_p$. The physical implications of these observations are that in a medium like a dielectric the group velocity is the velocity at which the maximum of the wave is propagating and so represents the velocity of propagation of energy. The phase velocity is the velocity of the wave front, this being the plane along which the phase has, at a given instant, the same value.

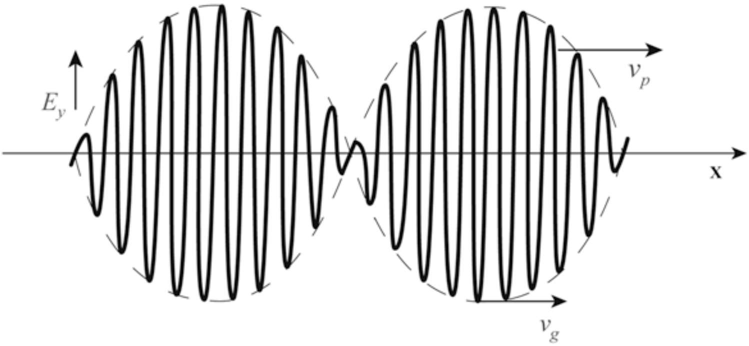


**Fig. 3.2** Group and phase velocity

### 3.2.2 The Appleton–Hartree Formula

After the previous introduction to the basic equations of an electromagnetic wave propagating in free space, wave propagation in a particular medium is now considered: plasma subject to an external magnetic field. This briefly explains how the concepts of the refraction index, given by the Appleton-Hartree formula, the gyrofrequency, and the plasma frequency can describe propagation in the ionosphere.

In a medium other than vacuum the resulting attenuation must also be considered. For this reason, it is preferable to express the equations in complex form (see Appendix A). For a simple case of linearly polarized wave, the Eq. 3.2.15 can be written as follows:

$$E_y = A\, exp\, i\,(\omega t\; - \;kx) \tag{3.2.21}$$

In which the decrease in amplitude $A$ could be accounted for with the function $exp\,(-ax)$. In fact here the coefficient of phase $k = \omega n/c$ defined above is now a complex quantity as well as the index of refraction $n = \mu - i\beta$ with $\mu$ the real part and $a = \omega\beta/c$. So, substituting these quantities in Eq. 3.2.21 and remembering that $i^2 = -1$:

$$E_y = A\, exp\;\, i\,(\omega t - \omega nx/c) = A\, exp\;\, i\,(\omega t - \omega\,(\mu - i\beta)\,x/c) \tag{3.2.22}$$

and:

$$E_y = A\;\, exp\, i\,(\omega t\; - \;\omega\mu x/c\; + \;i\omega\beta x/c) \tag{3.2.23}$$

Then similarly to Eq. 3.2.15:

$$E_y = A\sin\,(\omega t\; - \;kx)\;\, exp(-ax) \tag{3.2.24}$$

In this brief treatment of magneto-ionic theory, a series of simplifications need to be applied: the first is the nature of the medium of propagation, which should be considered as composed by ions, electrons, molecules, and be totally neutral. The external magnetic field should be almost constant or slowly variable in comparison to the frequency of the wave, and finally this magnetoplasma should be considered as cold so that thermal movements can be ignored.

When an electromagnetic wave propagates in such a medium, charged particles like ions and electrons are subject to the force applied by the electric field **E** of the wave. The vibration of these particles has an amplitude and phase directly related to their individual oscillation characteristics, returning some of the energy as radiation and losing some energy through collisions with surrounding particles. These collisions explain wave absorption while wave dispersion can be related to the differing oscillation frequencies of the individual particles.

If we consider the propagation of a plane harmonic wave in the nominally cold magnetoplasma of this simplified ionosphere, the interaction with positive or negative charged particles can be described with the following dynamics equation:

$$m d\mathbf{u}/dt = q\mathbf{E} + q\mathbf{u} \times \mathbf{B_0} - m\upsilon\mathbf{u} \tag{3.2.25}$$

where $\mathbf{u}$ is the velocity and $m$ the mass of the free electrons in the cold magnetoplasma, $q\mathbf{E}$ is the force applied by the electric field $\mathbf{E}$, $q\mathbf{u} \times \mathbf{B_0}$ the force produced by the vector of the magnetic induction when acting over a charge $q$ with velocity $\mathbf{u}$, according to Lorentz's law, and finally $m\upsilon\mathbf{u}$ is the viscosity force produced by the collisions proportional to the velocity and frequency $\upsilon$ of the collisions.

The average movement of electrons in a magnetoplasma can be described as a current with:

$$\mathbf{j} = Nq\mathbf{u} \tag{3.2.26}$$

where $N$ is the density of the electrons.

Rewriting $\mathbf{u}$ in a complex form: $\mathbf{u} = \mathbf{u}_0\, exp\, i\,(\omega t - \varphi)$ then:

$$d\mathbf{u}/dt = i\omega\mathbf{u} = i\omega\, \mathbf{j}/Nq \tag{3.2.27}$$

so Eq. 3.2.25 is:

$$m\, i\omega\mathbf{j}\,/Nq = q\mathbf{E} + q\mathbf{j}/Nq \times \mathbf{B_0} - m\upsilon\mathbf{j}/Nq \tag{3.2.28}$$

which dividing for $mi\omega/Nq$ and multiplying by $i^2$ gives:

$$-\mathbf{j} = iNq^2\mathbf{E}/m\omega + iq\mathbf{j}/m\omega \times \mathbf{B_0} - i\upsilon\mathbf{j}/\omega \tag{3.2.29}$$

Now if we define the following quantities:

$$X = f_N^2/f^2 \tag{3.2.30}$$

where $f_N^2 = Nq^2/4\pi^2\varepsilon_0 m$ and:

$$Y = f_H/f \tag{3.2.31}$$

with $f_H = q|\mathbf{B_0}|/2\pi m$ and:

$$Z = \upsilon/2\pi f \tag{3.2.32}$$

the Eq. 3.2.29 can be written as:

$$-\mathbf{j} = i\varepsilon_0\omega X\mathbf{E} + i\mathbf{j} \times \mathbf{Y} - iZ\mathbf{j} \tag{3.2.33}$$

It is important to underline the physical meaning of two quantities mentioned above, these being:

1. $f_N^2 = Nq^2/4\pi^2\varepsilon_0 m$ the plasma frequency, which represents the frequency of free oscillation of the electrons;
2. $f_H = q|\mathbf{B_0}|/2\pi m$ the gyrofrequency, which is the frequency of the circular motion of the free electrons subject to a magnetic induction field $\mathbf{B_0}$ according to Lorentz's law.

Furthermore, if the vector $\mathbf{B_0}$ is taken to be lying in the plane *xy,* so that $B_T$ is its component along the *y* axis, perpendicular to the direction of propagation of the wave, and $B_L$ its component along the *x* axis in the same direction of propagation. Accordingly:

$$Y_T = qB_T/\omega m = (q/\omega m)|\mathbf{B_0}|\sin\vartheta \tag{3.2.34}$$

$$Y_L = qB_L/\omega m = (q/\omega m)|\mathbf{B_0}|\cos\vartheta \tag{3.2.35}$$

with $\vartheta$ the angle between $\mathbf{B_0}$ and the axis $x$.

So the Eq. 3.2.33 along the 3 axes can be written as:

$$-j_x = i\varepsilon_0\omega XE_x + ij_zY_T - iZj_x \tag{3.2.36}$$

$$-j_y = i\varepsilon_0\omega XE_y + ij_zY_L - iZj_y \tag{3.2.37}$$

$$-j_z = i\varepsilon_0\omega XE_z + ij_xY_T - ij_yY_L - iZj_z \tag{3.2.38}$$

Remembering the last two Maxwell Eqs. 3.2.3 and 3.2.4 given above in relation to free space, these could now be rewritten for the case of ionospheric plasma as:

$$\text{curl }\mathbf{E} = -\delta\mathbf{B}/\delta t \tag{3.2.39}$$

$$\text{curl }\mathbf{B} = \mu_0\mathbf{j} + \varepsilon_0\mu_0\,\delta\mathbf{E}/\delta t \tag{3.2.40}$$

where $\mathbf{j} = \sigma\mathbf{E}$ is density of electronic current with $\sigma$ the conductivity. Applying the *curl* operator to the Eqs. 3.2.39 and 3.2.40, it is possible to obtain by mathematical steps, listed in Appendix B, the following equations:

$$j_x = -i\omega\,\varepsilon_0 E_x \tag{3.2.41}$$

$$j_y = i\omega\,\varepsilon_0\left(n^2 - 1\right)E_y \tag{3.2.42}$$

$$j_z = i\omega\,\varepsilon_0\left(n^2 - 1\right)E_z \tag{3.2.43}$$

From these equations and from the Eqs. 3.2.36, 3.2.37, and 3.2.38, it is possible to deduce the Appleton-Hartree equation for the complex index of refraction $n$ (see the complete mathematical passages in the Appendix B):

$$n^2 = (\mu - i\beta)^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \sqrt{\left(\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_L^2\right)}} \tag{3.2.44}$$

The double index of refraction $n$ means that when an electromagnetic wave propagates in a dispersive and anisotropic medium, the propagation of two waves of the same frequency and with different phase velocities are observed. These two waves, the magneto-ionic components, are not observable in free space, but can be identified when they propagate differently in an anisotropic medium like magnetoplasma.

Now, if it is assumed that in the ionosphere the frequency of collisions is negligible in comparison with the frequency of the wave, so that $Z = 0$, then the real part of the Appleton-Hartree equation is:

$$n^2 = 1 - \frac{2X(1 - X)}{2(1 - X) - Y_T^2 \pm \sqrt{Y_T^4 + 4(1 - X)^2 Y_L^2}} \tag{3.2.45}$$

Furthermore, in the absence of an external magnetic field $Y_T = Y_L = 0$ the Appleton-Hartree equation for the phase refractive index can be written in its simplest form:

$$n^2 = 1 - X = 1 - f_N^2/f^2 \tag{3.2.46}$$

To complete this discussion of magneto-ionic theory, other approximations can be made significantly simplifying the Appleton-Hartree equation. If the propagation is almost perpendicular to the direction of the magnetic field, then:

$$Y_T^4 \gg 4(1 - X - iZ)^2 Y_L^2, \tag{3.2.47}$$

or almost parallel to the magnetic field, then:

$$Y_T^4 \ll 4(1 - X - iZ)^2 Y_L^2 \tag{3.2.48}$$

In order to establish simple relations that describe the penetration and total reflection of an electromagnetic wave, as in the case of vertical ionospheric radio sounding, the simple assumptions considered above regarding propagation in a homogenous magnetoplasma with $Z = 0$ are reintroduced.

If for a total reflection $n = 0$, and assuming above $Z = 0$ and $\beta = 0$, since absorption is considered negligible, the Appleton-Hartree equation could be written as:

$$1 - X = 0 \tag{3.2.49}$$

$$(1 - X)^2 = Y_T^2 + Y_L^2 = Y^2 \tag{3.2.50}$$

From these equations, substituting Eqs. 3.2.30 and 3.2.31 in $X$ and $Y$, the first can be solved as:

$$f = f_N \tag{3.2.51}$$

and the second, as:

$$\left(1 - f_N^2/f^2\right)^2 = f_H^2/f^2 \tag{3.2.52}$$

Then by solving a second order equation the following relations are obtained for $f_z$ and $f_x$

$$f_z = \sqrt{\left(f_N^2 + \frac{1}{4}f_H^2\right)} - \frac{1}{2}f_H \tag{3.2.53}$$

$$f_x = \sqrt{\left(f_N^2 + \frac{1}{4}f_H^2\right)} + \frac{1}{2}f_H \tag{3.2.54}$$

In the case of the ionospheric layers, assuming an electron density increasing with altitude to a maximum electron density $N_M$ and $f_H \leq 2f_N$, the previous equations could be rewritten in the approximate forms of:

$$f_x = f_{NM} + f_H/2 \tag{3.2.55}$$

$$f_z = f_{NM} - f_H/2 \tag{3.2.56}$$

$$f_z = f_{NM} - f_H/2 \tag{3.2.57}$$

These are the three critical frequencies that can be observed in an ionogram produced by ionospheric vertical sounding.

## 3.3 Vertical and Oblique Incidence Sounding in the Ionosphere

### *3.3.1 Vertical Incidence Sounding*

As mentioned in the historical notes in the Introduction, the traditional method for ionospheric measurements is vertical ionospheric sounding. A dense network of instruments for these measurements around the world produce the most widely available routine observations, organized following standard rules. They have contributed to the development of a general knowledge of the morphology of the ionosphere (see Chap. 4), used for frequency planning of long distance HF radio communications. The experimental technique first initiated by G. Breit and M. A. Tuve in 1925 can be considered the precursor of the radar technique.

The method and instrumentation used to measure the electron density of the ionosphere as a function of altitude is based on the principle that when an electromagnetic wave vertically penetrates ionospheric plasma, reflection occurs at the level where the refractive index $n$ becomes zero.

According to the magneto-ionic theory the refractive index:

$$n^2 = 1 - (f_N/f)^2 \tag{3.3.1}$$

is dependent on the plasma frequency $f_N$ given by:

$$f_N = \sqrt{\frac{Nq^2}{4\pi^2 \varepsilon_0 m}} \tag{3.3.2}$$

where $N$ is electron density, $q$ and $m$ the charge and the mass of the electron, $\varepsilon_0$ the dielectric constant in vacuum, and $f$ the incident frequency. Reflection in the ionosphere occurs when the incident frequency $f$ is equal to $f_N$.

A vertical ionospheric sounder (ionosonde), designed to detect electron density of ionospheric plasma as a function of height, emits radio impulses scanning the transmitting frequency from 1 to 20 MHz, and measures the time delay of any echoes (see Fig. 3.3).

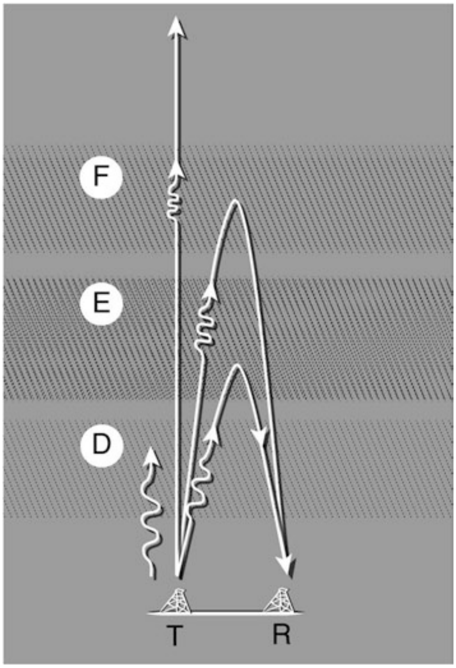


**Fig. 3.3** A vertical ionospheric sounder emits radio impulses with increasing frequency from 1 to 20 MHz, measuring the time delay of radio signals received back from the different ionospheric layers

As the incident frequency increases the radio impulses penetrate the ionosphere up to higher altitudes and are reflected by higher levels of electron density $N$. However, according to the electron density profile shown in Fig. 2.21, the electron density $N$ varies and increases with altitude up to a relative maximum and even an absolute maximum $N_{M.}$ The maximum electron density $N_M$ corresponds to the maximum reflected incidence frequency, called the critical frequency $fo$ as in Eq. 3.2.56:

$$N_M = 1.24\, 10^{10} fo^2 \tag{3.3.3}$$

where $N_M$ and $fo$ are expressed in el/m$^3$ and in MHz respectively.

It is important to note that the critical frequency $fo$ of a ionospheric layer linked to the relative maxima for the E region ($foE$), the F1 layer ($foF1$), and the absolute maximum for the F2 layer ($foF2$), are referred only to the ordinary ray. If extraordinary rays are also considered, the maximum frequencies reflected, corresponding to the other magneto-ionic components, are:

$$fx = f_{NM} + f_H/2 \tag{3.3.4}$$

$$fz = f_{NM} - f_H/2 \tag{3.3.5}$$

where $f_{NM} = fo$ is the critical frequency of the ordinary ray, $fx$ the critical frequency of the extraordinary ray, $fz$ the critical frequency of the $z$ ray, and $f_H$ the known gyrofrequency.

The height of the point where the incident frequency is reflected is given by:

$$h' = c\Delta t/2 \tag{3.3.6}$$

where $c$ is the velocity of the electromagnetic pulse in free space and $\Delta t$ is the delay between the transmission time and the received echo.

It is important to note that $h'$ is not the real height of reflection but is defined as the virtual height or group height. Considering that an electromagnetic pulse travels through a plasma at lower velocity than light in a vacuum, the time delay is:

$$\Delta t = \frac{2}{c}\int_0^h \mu' dh = \frac{2}{c}\int_0^h dh/\mu = \frac{2}{c}h' \tag{3.3.7}$$

where $\mu'$ is the group refractive index and $\mu$ the refractive index of the phase that, as seen in the Sect. 3.2, are functions of electron density $N$, frequency of the incident radio pulse $f$, and gyrofrequency $f_H$. Consequently, the height measurement is virtual, this being greater than the true height and becoming much greater as the electromagnetic pulse is slowed as the ionosphere approaches maximum electron density. Finally, when the electromagnetic pulse is equal to or exceeds the maximum density the time delay is infinite, and so is the virtual height.

### 3.3.2 *Oblique and Backscatter Sounding*

While in vertical ionospheric sounding the transmitter is in the same place or very close to the receiver, in oblique sounding the receiver is located at a considerable distance, even thousands of kilometres, with multiple reflections from the ionosphere or off the ground (Fig. 3.4).

Again in this case, the transmitter emits radio impulses of increasing frequency from 1 to over 20 MHz because the secant law (see Chap. 9) allows reflections in the ionosphere of higher frequencies. The echoes are received at the distant receiver after a given time delay and with opportune synchronization of the two devices. The traditional synchronization method, still widely used today, is the chirp technique, with more recently the GPS clock method.

The result is an oblique ionogram in which it is possible to directly read the Maximum Observable Frequency (*MOF*), for the given distance, making these measurements extremely important for testing any long or short-term predictions for HF frequency parameters (see Fig. 3.5 and also Chap. 10).

The oblique ionogram shows not only the maximum frequency reflected for that distance but also double delays for the same frequency when the upper part of the trace seems to come back, suggesting the possibility of a double radio path.

Another oblique sounding method is backscattering, in which radio impulses transmitted from a very high power ionosonde are first reflected by the ionosphere,

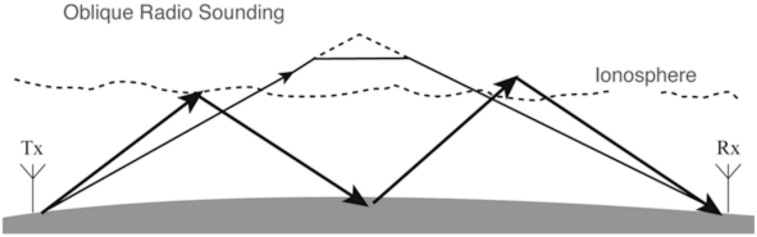


**Fig. 3.4** An oblique radio sounding setup between two points Tx and Rx

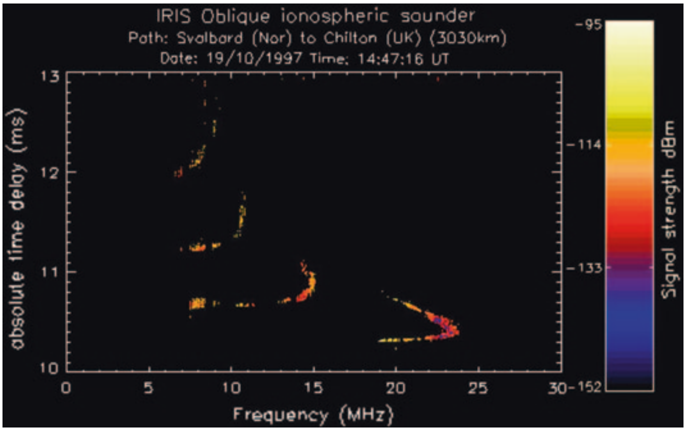


**Fig. 3.5** Oblique ionogram between points in Svalbard (Norway) and Chilton (UK) 3,030 km apart (Courtesy Rutherford Appleton Laboratory)

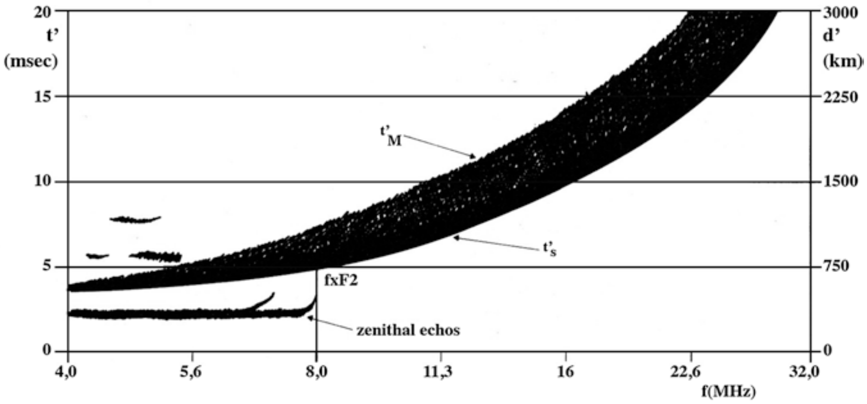


**Fig. 3.6** Example of a backscatter ionogram (from Dominici 1971)

then scattered by the ground, then finally return to the equipment travelling on the same path. The result is a backscatter ionogram (Fig. 3.6) used to detect the skip distance area or to monitor the quiet and disturbed ionosphere over a given region around the transmitter point. However, the high power required and the large and complex arrays of wide band rhombic antenna with very low radiation angles, results in these experiments being rather limited in number.

## 3.3.3 *Ionosondes*

### 3.3.3.1 Main Characteristics

Since the first years of ionospheric research, when only fixed radio frequencies were used, important improvements and enormous progress has been achieved. However, the technical principle of the most widely used method for systematic measurement of the electron density in the upper atmosphere has remained substantially the same. The device used is called an ionosonde and can be described as a bistatic radar system. This means that the transmitting antenna is separated from the receiving antenna and that there are two separate energy paths. Figure 3.7 schematically illustrates the components of an ionosonde:

1. a synthesizer that generates the radio frequency to be transmitted and a transmitter that amplifies the electromagnetic pulse to be emitted from an antenna system;
2. a receiver *apparatus* with an antenna system that amplifies the received signal;
3. finally, a control system able to pilot the emission time in accordance with the expected arrival of the signal at the receiver, and analyze and elaborate the signal received as a graphic trace.

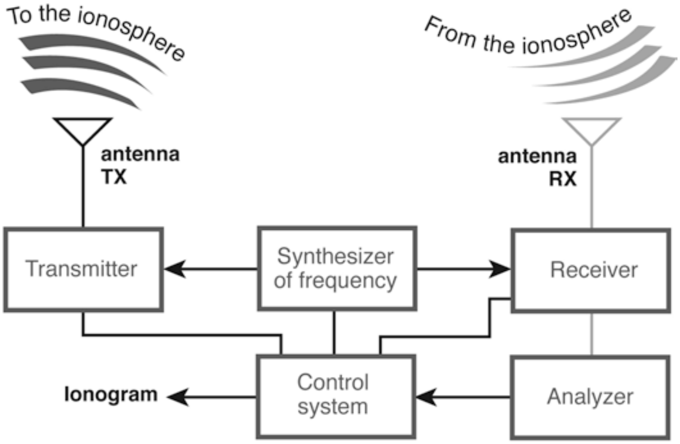


**Fig. 3.7** Simple scheme of the main components of a vertical ionospheric sounder

The ionosonde can also be interfaced to peripheral devices like video screens, plotters or computers for further analysis or autoscaling. Over the last century, these experimental instruments have evolved significantly, often driven by the scientific priorities of the inventors and consequently many different types of homemade ionosondes were produced in the laboratories of different research institutes. Recently, the spread of newer commercial ionosondes ensures the widespread availability of high quality observations from many locations around the world.

The first *archeo*-ionosondes before the Second World War employed transmission and receiving techniques that were only partly electronic and required complicated tuning mechanisms. Since then the most important development in ionosonde devices include the transition from analogical to digital techniques in the 1970s, from envelope to compression impulses in the 1990s, and most recently the development of robust autoscaling software applications.

The transition from envelope to compression impulses was particularly important because it significantly improved the performance of the new ionosondes. In the original radar technique using the envelope principle, the pulses transmitted were long $\tau$ and the received signals, after a time $\Delta t$, over the envelope pulse were still long $\tau$ as shown in the upper and lower traces of Fig. 3.8, left. Therefore, an evaluation of the spatial resolution of the system, which requires distinguishing two different signals, can only be achieved when the difference in the receiving times $t2$ and $t1$ is greater than the time $\tau$, $t2 - t1 \geq \tau$ (Fig. 3.8, right.). Thus the minimum distance $\delta d$ is:

$$\delta d = c\tau/2 \tag{3.3.8}$$

Instead, in the new technique using the so-called pulse compression within a long pulse of length $\tau$ there is a modulation of the phase of length $t_s$ with the phase remaining constant. The sequence of phase changes is codified and analyzed in

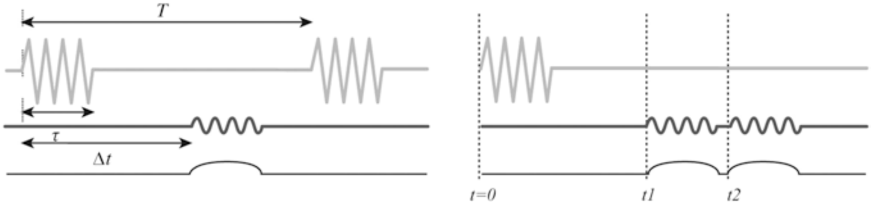


**Fig. 3.8** Spatial resolution of the envelope technique used in the old original ionosondes: on the *left* the *upper graph* represents the successive transmission of two signals of length $\tau$ received, *second line* in the figure, after a delay of $\Delta t$. On the *right*, it is shown how two signals cannot be distinguished if the differences in the times of arrival $t1$ and $t2$ are less than $\tau$

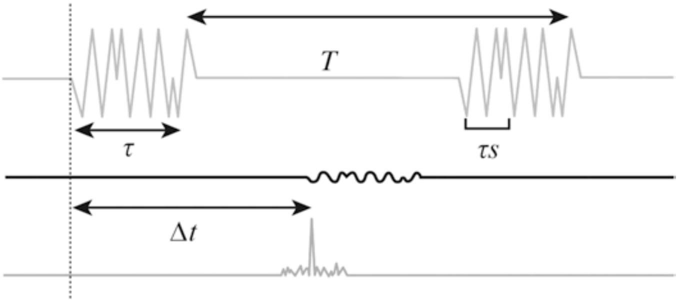


**Fig. 3.9** The *upper line* represents the successive transmission of two signals of length $\tau$ with a short modulation of the phase of length $\tau_s$, this modulation is codified in order to emphasize the received signal and greatly reduce the limit of spatial resolution

order to emphasize the received signal (Fig. 3.9). In this case, the spatial resolution becomes $\delta d = c\tau_s/2$, which is much less than the previous system while still maintaining an extended duration of pulse $\tau$.

The typical characteristics and advantage of the two systems can be summarized in the following Table 3.1.

**Table 3.1** Typical characteristics of the two systems

| | Envelope pulses | Compression pulses |
|---|---|---|
| Frequency range | 1–20 MHz | 1–20 MHz |
| Height range | 90–1,000 km | 90–1,000 km |
| Pulse amplitude | 100 μs | 480 μs (sub pulse 30 μs) |
| Peak power TX | 1–10 kW | 250 W |
| Height resolution | 15 km | 5 km |
| Repetition rate | 100 Hz | 60 Hz |
| Reception | Analogue | Digital |
| Output (ionogram) | Paper or film | File |

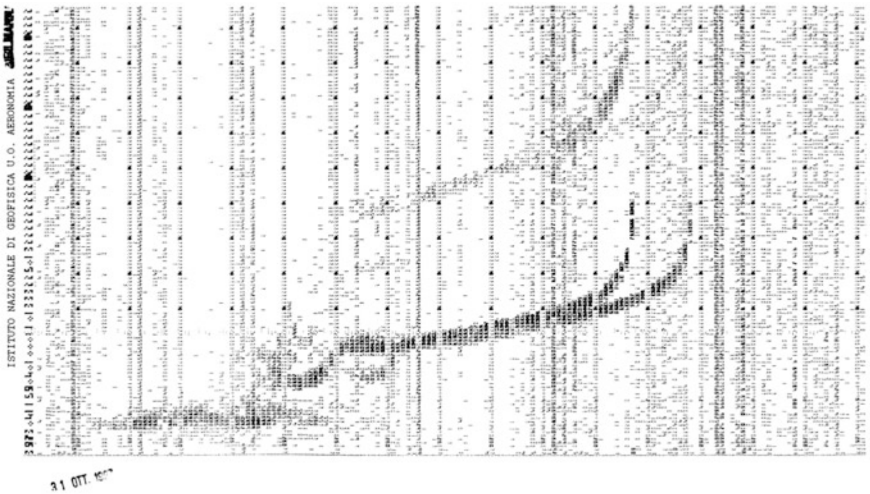


**Fig. 3.10** A digital ionogram from a digisonde DGS 256 operating at the Rome (41.9° N, 12.5° E) ionosonde station: the trace recorded on paper provides digital information on the strength of the received signal

#### 3.3.3.2 The Digital Ionosonde or Digisonde

In the late 1960s, the technical revolution of applied solid state electronics and the progress of computer systems enabled the development of digital ionosondes at Lowell University, first the 128 P, then in the 1980s the DGS 256, and finally from the 1990s and continuing now the most up to date versions of the Digisonde Portable Sounder DPS 4. The capabilities of these devices are extremely important not only for digitally tracking ionograms (Fig. 3.10) with the recording of ionograms first on tape and then on firmware, but also for the possibility of new measurements, like the possibility of separating the ordinary from the extraordinary ray, or drift measurements obtained using a complex receiving array antenna system able to measure the Doppler shifts of signals arriving from different directions. The latest generations of ionosondes are also programmable and, using a routine for automatic interpretation of ionograms, they can be interfaced with the Internet for space weather applications.

#### 3.3.3.3 Antenna System

To complete this brief description of ionosonde *apparatus,* it is necessary to mention the antenna systems used. Considering the range of frequencies used by the ionosonde, 1–20 MHz, wideband, log periodic delta, or rhombic antennas must be used. At vertical sounding stations rhombic or double delta vertical arrays are generally used, positioned above perpendicular planes, one for transmitting and the second for receiving signals (Fig. 3.11, left). They terminate with a load resistor

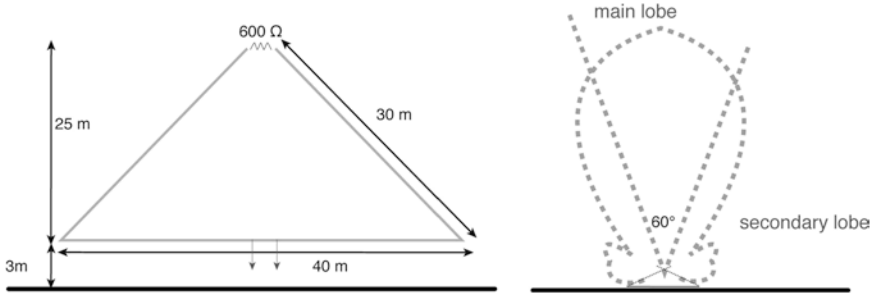


**Fig. 3.11** On the *left,* a simple scheme and dimensions of a transmitting delta array with a 600 Ω resistor at the *top*, and on the *right,* the graph of directional gain

which raises the directional gain toward the end, giving good response over a wider frequency range (Fig. 3.11, right).

### *3.3.4 Ionograms and Their Interpretation*

The routine tasks of every standard ionospheric station, as defined in the URSI Handbook of ionogram interpretation and reduction edited by W. R. Piggot and K. Rawer, are:

- Monitor the ionosphere above the station.
- Obtain significant median data to evaluate long-term changes.
- Study phenomena peculiar to the region.
- Study the global morphology of the ionosphere.

These objectives require a set of standard techniques and conventions applicable for the general interpretation of ionospheric measurements in order to achieve a more phenomenological description of the ionogram, as well as providing a simplified description of the ionosphere above the station. This set of conventions and rules for scaling ionograms are exhaustively described in the handbook mentioned above and more recently in the Manual of ionogram scaling, edited by the Japanese Ministry of Post and Telecommunications, Radio Research Laboratory. For a simple description of the ionosphere by vertical sounding it is convenient to consider the ionosphere as schematically divided into the conventional regions D, E, F, and the sporadic Es. Figure 3.12 shows an ideal daytime ionogram with ordinary and extraordinary traces and their specific characteristics.

**F Region**

*foF2*: critical frequency of the ordinary trace of the highest layer of the F region, called the F2 layer when the F1 layer is present.
*fxI*: highest frequency recorded by a reflection from the F region.
*foF1*: critical frequency of the ordinary trace of the F1 layer when present.

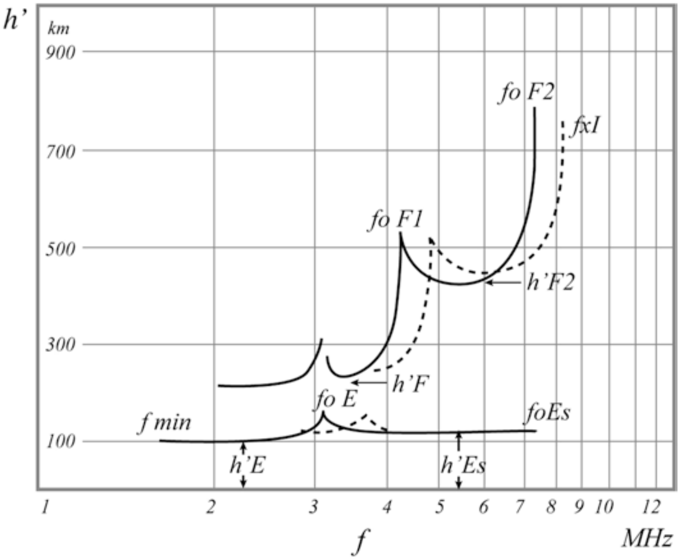


**Fig. 3.12** Ideal daytime ionogram with routinely scaled characteristics

*h′F2*: minimum virtual height of the ordinary trace of the F2 layer.
*h′F*: lowest virtual height of the ordinary trace of the F region.

### E Region

*foE*: critical frequency of the ordinary trace of the E region.
*h′E*: minimum virtual height of the ordinary trace of the E region.

### Es Sporadic E layer

*foEs*: highest frequency of the ordinary trace of the continuous sporadic E layer.
*h′Es*: minimum virtual height of the ordinary trace of the Es layer.
*fbEs*: blanketing frequency of the Es layer.
*fmin*: lowest frequency recorded in the ionogram.

Another characteristic that is routinely scaled at a standard ionospheric station, and which plays an important role in HF radio propagation prediction, is the transmission factor *M* or *M(3000)F2*, as will be described in detail in Chap. 9. The *M* factor is a conversion factor to obtain the Maximum Usable Frequency over a given oblique propagation distance. The *M* factor for a standard distance of 3,000 km is referred to as *M(3000)F2* and is related to the associated Maximum Usable Frequency *MUF(3000)F2* and to the critical frequency *foF2* in:

$$M(3000)F2 = MUF(3000)F2/foF2 \tag{3.3.9}$$

In the case of a reflection above the F1 layer the *M* factor can be considered as *M(3000)F1*, not usually a routinely scaled parameter and also related to the critical frequency.

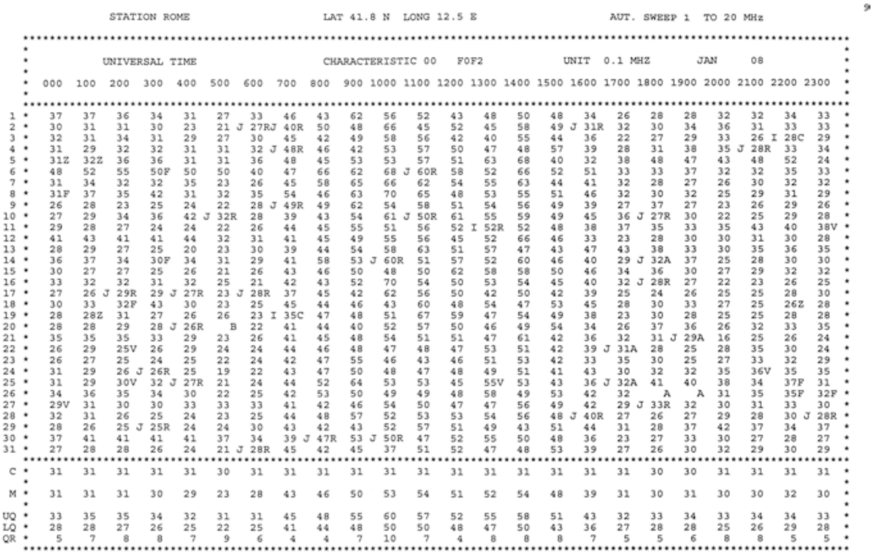
### 3.3.4.1 Presentation and Processing of Ionospheric Data

The measurements of the principal ionospheric characteristics obtained from scaled ionograms are reported in daily and monthly tables. For each characteristic, two letters, the first qualifying and the second descriptive, help provide a complete interpretation and description of the ionograms. The monthly tables of hourly ionospheric data also include the computed values for the monthly median and the upper and lower quartiles (Fig. 3.13).

In the past, these tables constituted the monthly bulletin of vertical ionospheric sounding, communicated to World Data Centers and other scientific institutions. Today they contribute to ionospheric databases on the Internet, like the following examples:

1. *SPIDR Space Physics Interactive Data Resource:* http://spidr.ngdc.noaa.gov/spidr/
2. *ESWUA Electronic Space Weather Upper Atmosphere:* http://www.eswua.ingv.it/
3. *DIAS Digital Ionospheric Upper Atmosphere Service:* http://www.iono.noaa.gr/DIAS/
4. *UMass. Lowell Center for Atmospheric Research:* http://umlcar.uml.edu/.

The monthly median value for a given hour is defined as the middle value of a set of values, at the same hour, arranged in order of magnitude. Similarly, the upper and lower quartile represent the upper half and the lower half of the same hourly values still arranged in order of magnitude. As will be seen in the following



```
            STATION ROME                      LAT 41.8 N  LONG 12.5 E                 AUT. SWEEP 1  TO 20 MHz

    *************************************************************************************************************************
    *                                                                                                                       *
    *          UNIVERSAL TIME                     CHARACTERISTIC 00   FOF2            UNIT  0.1 MHZ       JAN      08       *
    *                                                                                                                       *
    *  000  100  200  300  400  500  600  700  800  900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 *
    *                                                                                                                       *
    *************************************************************************************************************************
  1 *   37   37   36   34   31   27   33   46   43   62   56   52   43   48   50   48   34   26   28   28   32   32   34   33 *
  2 *   30   31   31   30   23   21 J 27RJ 40R  50   48   66   45   52   45   58   49 J 31R  32   30   34   36   31   33   33 *
  3 *   32   31   34   31   29   27   30   45   42   49   58   56   42   40   55   44   36   22   27   29   33   26 I 28C  29 *
  4 *   31   29   32   32   31   31   32 J 48R  46   42   53   57   50   47   48   57   39   28   31   38   35 J 28R  33   34 *
  5 *   31Z  32Z  36   36   31   31   36   48   45   53   53   57   51   63   68   40   32   38   48   47   43   48   52   24 *
  6 *   48   52   55   50F  50   50   40   47   66   62   68 J 60R  58   52   66   52   51   33   33   37   32   32   35   33 *
  7 *   31   34   32   32   35   23   26   45   58   65   66   62   54   55   63   44   41   32   28   27   26   30   32   32 *
  8 *   31F  37   35   42   31   32   35   54   46   63   70   65   48   53   55   51   46   32   30   32   25   29   31   29 *
  9 *   26   28   23   25   24   22   28 J 49R  49   62   54   58   51   54   56   49   39   27   37   27   23   26   29   26 *
 10 *   27   29   34   36   42 J 32R  28   39   43   54   61 J 50R  61   55   59   49   45   36 J 27R  30   22   25   29   28 *
 11 *   29   28   27   24   24   22   26   44   45   55   51   56   52 I 52R  52   48   38   37   35   33   35   43   40   38V*
 12 *   41   43   41   41   44   32   31   41   45   49   55   56   45   52   66   46   33   23   28   30   30   31   30   28 *
 13 *   28   29   27   25   20   23   30   39   44   54   58   63   51   57   47   43   47   43   38   33   30   35   36   35 *
 14 *   36   37   34   30F  34   31   29   41   58   53 J 60R  51   57   52   60   46   40   29 J 32A  37   25   28   30   30 *
 15 *   30   27   27   25   26   21   26   43   46   50   48   50   62   58   58   50   46   34   36   30   27   29   32   32 *
 16 *   33   32   32   31   32   25   21   42   43   52   70   54   50   53   54   45   40   32 J 28R  27   22   23   26   25 *
 17 *   27   26 J 29R  29 J 27R  23 J 28R  37   45   42   62   56   50   42   50   42   39   25   24   26   25   25   28   30 *
 18 *   30   33   32F  43   30   23   25   45   44   46   43   60   48   54   47   53   45   28   30   33   27   25   26Z  28 *
 19 *   28   28Z  31   27   26   26   23 I 35C  47   48   51   67   59   47   54   49   38   23   30   28   25   25   28   28 *
 20 *   28   28   29   28 J 26R    B  22   41   44   40   52   57   50   46   49   54   34   26   37   36   26   32   33   35 *
 21 *   35   35   35   33   29   23   26   41   45   48   54   51   51   47   61   42   36   32   31 J 29A  16   25   26   24 *
 22 *   26   29   25V  26   29   24   24   44   46   48   47   48   47   53   51   42   39 J 31A  28   25   28   35   30   24 *
 23 *   26   27   25   24   25   22   24   42   47   55   46   43   46   51   53   42   33   35   30   25   27   33   32   29 *
 24 *   31   29   26 J 26R  25   19   22   43   47   50   48   47   48   49   51   41   43   30   32   32   35   36V  35   35 *
 25 *   31   29   30V  32 J 27R  21   24   44   52   64   53   53   45   55V  53   43   36 J 32A  41   40   38   34   37F  31 *
 26 *   34   36   35   34   30   22   25   42   53   50   49   49   48   58   49   53   42   32     A    A  31   35   35F  32F*
 27 *   29V  31   30   30   33   33   33   41   42   46   54   50   47   47   56   49   42   29 J 33R  32   30   31   33   30 *
 28 *   32   31   26   25   24   23   25   44   48   57   52   53   53   54   56   48 J 40R  27   26   27   29   28   30 J 28R*
 29 *   28   26   25 J 25R  24   24   30   43   42   43   52   57   51   49   43   51   44   31   28   37   42   37   34   37 *
 30 *   37   41   41   41   41   37   34   39 J 47R  53 J 50R  47   52   55   50   48   36   23   27   33   30   27   28   27 *
 31 *   27   28   28   26   24   21 J 28R  45   42   45   37   51   52   47   48   53   39   27   26   30   32   29   30   29 *
    *************************************************************************************************************************
  C *   31   31   31   31   31   30   31   31   31   31   31   31   31   31   31   31   31   31   30   30   31   31   31   31 *
    *                                                                                                                       *
  M *   31   31   31   30   29   23   28   43   46   50   53   54   51   52   54   48   39   31   30   31   30   30   32   30 *
    *                                                                                                                       *
 UQ *   33   35   35   34   32   31   31   45   48   55   60   57   52   55   58   51   43   32   33   34   33   34   34   33 *
 LQ *   28   28   27   26   25   22   25   41   44   48   50   50   48   47   50   43   36   27   28   28   25   26   29   28 *
 QR *    5    7    8    8    7    9    6    4    4    7   10    7    4    8    8    8    7    5    5    6    8    8    5    5 *
    *************************************************************************************************************************
```

**Fig. 3.13** A page of a typical monthly ionospheric bulletin issued by an ionospheric station according to the URSI rules of scaling

Chapters, these statistical parameters have great importance for the climatological models of the ionosphere and consequently for long term prediction of HF radio propagation conditions. However, monthly median values cannot be considered as typical of quiet conditions of the ionosphere. They are obviously dependent on the number of values included in the set.

#### 3.3.4.2 Autoscaling

For much of the history of ionospheric observation the interpretation of ionospheric recordings, the so-called scaling of ionograms, was an enormous and expensive job in terms of skills and manpower. The evolution of digital electronic digisondes made it possible to evaluate an ionogram trace for signal amplitude together with polarization between ordinary and extraordinary rays, as well as distinguishing vertical from off-vertical echoes. On the basis of these advances, at the beginning of the 1980s B.W. Reinish and X. Huang proposed an operating method for complete automatic scaling of ionograms and automatic calculation of electron density profiles, called ARTIST.

From the auto scaled *h′*(*f*) points a profile inversion algorithm is used to obtain the real electron density profile. Different models have been used for the different layers, for example parabolic for the E layer and valley, and a Chebyshev polinom for the F region. Improved versions of ARTIST in recent years have significantly enhanced the performance quality, reducing interpretation errors. Figure 3.14 shows an example ionogram with autoscaled values of the principal ionospheric characteristics, and the true electron density profile superimposed.

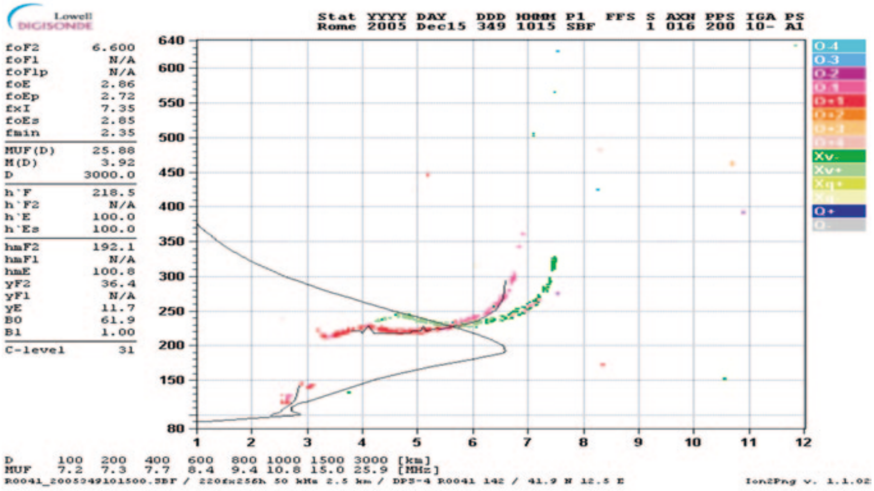
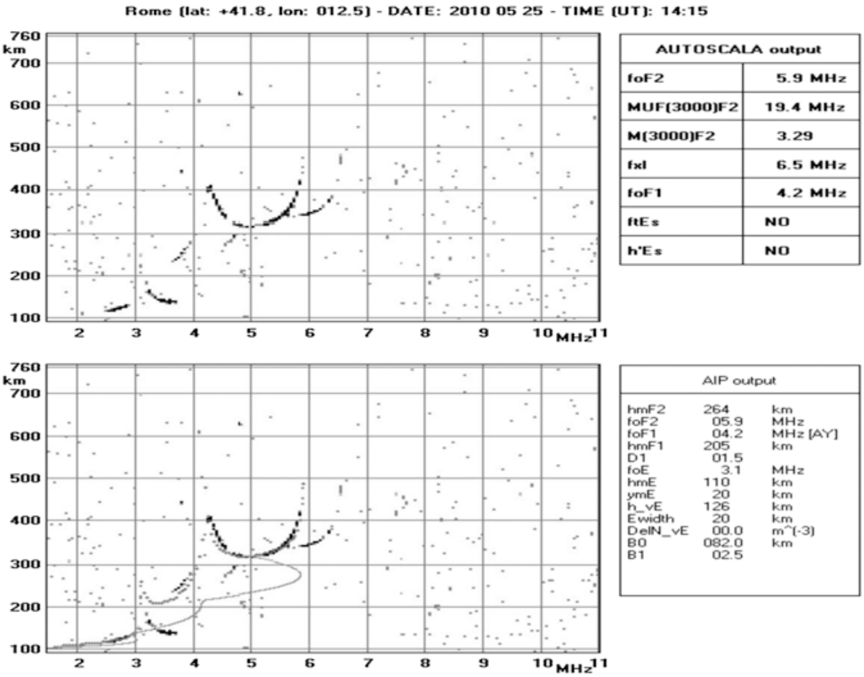


**Fig. 3.14** A DPS4 digital ionogram recorded at the Rome (41.9° N, 12.5° E) ionosonde station with autoscaled values of ionospheric characteristics on the *left* and the true electron density profile superimposed over the ionogram graph



| AUTOSCALA output | |
|---|---|
| foF2 | 5.9 MHz |
| MUF(3000)F2 | 19.4 MHz |
| M(3000)F2 | 3.29 |
| fxI | 6.5 MHz |
| foF1 | 4.2 MHz |
| ftEs | NO |
| h'Es | NO |

AIP output

| | | |
|---|---|---|
| hmF2 | 264 | km |
| foF2 | 05.9 | MHz |
| foF1 | 04.2 | MHz [AY] |
| hmF1 | 205 | km |
| D1 | 01.5 | |
| foE | 3.1 | MHz |
| hmE | 110 | km |
| ymE | 20 | km |
| h_vE | 126 | km |
| Ewidth | 20 | km |
| DelN_vE | 00.0 | m^(-3) |
| B0 | 082.0 | km |
| B1 | 02.5 | |

**Fig. 3.15** An example of the AUTOSCALA procedure: *above* is the original trace of an AIS ionosonde, on the *right* the autoscaled values of ionospheric characteristics, and *below* the true electron density profile

In addition to the possibility of automatically reading and archiving ionospheric observations, the autoscaling procedure has become extremely important for its capability for data assimilation and for nowcasting of ionospheric conditions, which is absolutely essential for operations affected by space weather. More recently, another autoscaling procedure has been developed by an Italian team, differing substantially from the one described above. The AUTOSCALA procedure is based on image recognition techniques. It does not require wave polarization data, and therefore, can be applied to any kind of ionosonde. The electron density profile is obtained by iteration, and starting from autoscaled anchor points applied to a model, it is able to reconstruct simulated ionograms and choose the one that best fits the real ionogram and so the nearest electron density profile (Fig. 3.15).

## 3.4 Ionospheric Scattering

Up to now the ionospheric medium has been treated as cold magnetized plasma (Sect. 3.2) in which thermal motions of electrons can be neglected. But when the real temperature of plasma is taken into account both radio and plasma waves can propagate in this medium. The fact that electron plasma waves are subject to marked attenuation and therefore do not travel over long distances has a significant consequence. They cannot be used for communication and consequently in most radio contexts they are justifiably ignored.

However, this is not the case for the theory of ionospheric scattering and its practical application in several types of scatter radar techniques currently used for ionospheric measurements. Obviously ionospheric scattering, whether incoherent or coherent, caused by thermal fluctuations, is so weak that it can be performed only with powerful transmitters, highly directive antennas and very sensitive receivers. Since a detailed description of warm plasma physics is beyond the scope of this book, only a summary is provided here of the main ionospheric scattering facilities, which produce very useful results for verifying and improving prediction and forecasting of ionospheric propagation.

### 3.4.1 Incoherent Scatter Radar: EISCAT

W. E. Gordon's idea in 1958 to use incoherent scatter (Thomson scattering) of radio waves on free electrons for measuring ionospheric concentration was experimentally proven for the first time by receiving continuously scattered waves at a frequency of 41 MHz in the initial observations by K. L. Bowles in the same year. The frequency of the high power radio beam sent into the ionosphere from a ground transmitter was much higher than the maximum plasma frequency in the ionosphere ($\omega \gg \omega_{pmax}$). It enabled an incoherent scatter radar (ISR) to measure a density profile up to the altitudes limited only by radar sensitivity, unlike ground-based ionosondes that measure ionospheric parameters up to the electron density maximum.

After this first significant breakthrough for radio science in general and ionospheric plasma ground-based measurements in particular, the incoherent scatter technique has rapidly been developed into a very powerful tool for remotely probing the Earth's ionosphere. It has been discovered that the observed scattering regions can be selected by adjusting the highly directive transmitting and receiving aerials, or by time gating the receiver. Since the received scattered signal is usually very weak, sophisticated signal integration methods have to be implemented in an extremely sensitive receiving device. The analysis of the scattered signal provides profiles of electron density, electron and ion temperatures, and ion velocity along the magnetic field lines typically from 80 to 1,000 km (Fig. 3.16).

The few well known ISR facilities in the world include the 49.92 MHz radar at the Jicamarca Radio Observatory (http://jicamarca.ece.cornell.edu/), a location very near the magnetic equator, the 430 MHz radar at Arecibo Observatory (http://www.naic.edu/), the 440 MHz Millstone Hill Observatory (http://www.haystack.edu/atm/mho/index.html), and the EISCAT (European Incoherent SCATter) Scientific Association (http://www.eiscat.se/) who operates three geophysical research Incoherent Scatter Radar systems at both the UHF (931 and 500 MHz) and VHF (224 MHz) regimes in northern Scandinavia (Fig. 3.17).

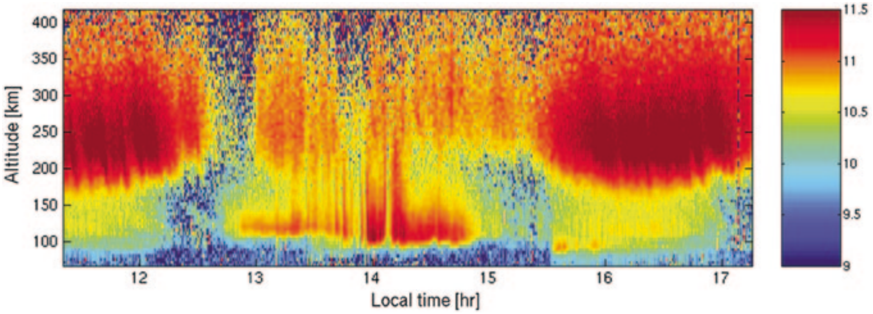


**Fig. 3.16** Decimal logarithm of electron density (electrons/m$^3$) measured by the EISCAT radar during the 9 February 1999 experiment (from Lilensten and Cander 2003)

**Fig. 3.17** Photo of the Svalbard 500 MHz radar system (http://www.eiscat.se/)

In addition to the incoherent scatter radar systems, EISCAT also operate an Ionospheric Heater facility at Ramfjordmoen and a dynasonde at Tromso. This combination of measurement facilities support various active plasma physics experiments and is thus contributing to a large expansion in knowledge of the high-latitude ionosphere and its polar cap region.

### *3.4.2 Coherent Scatter Radar: SuperDARN*

As described in the previous Section, the first ionospheric probing by radar observed backscattered signals at frequencies significantly above the maximum plasma frequency. In the case of coherent scatter radar, the receiver detects energy scattered within a medium when there are regular spatial variations in the refractive index due to ionosphere irregularities. Among the various coherent scatter radar facilities currently operating around the world is the Super Dual Auroral Radar Network (SuperDARN) (http://superdarn.jhuapl.edu/), which began formal operation in 1995. It is now a global-scale network of HF and VHF radar systems capable of sensing backscatter from irregularities in the E and F regions of the high-latitude ionosphere. It was derived from the DARN concept but is significantly larger in scope, comprising 14 radar stations in the northern hemisphere and 7 in the southern hemisphere. The essentially identical radar systems transmit a short sequence of pulses in the HF band in the frequency range from 8 to 20 MHz, and then sample the returning echoes. They operate continuously with a temporal resolution of 1–2 min. The final output of the processing procedure for this type of measurement is the calculated backscattered power, spectral width, and Doppler velocity of the plasma density irregularities in the ionosphere.

One of the most important applications of SuperDARN observations is the use of its high-latitude electric potential patterns as input data for the advanced ionospheric models, as well as assisting interpretation of tomography images over the area covered. The reconstruction of convection patterns using SuperDARN data is based on finding a functional form for the distribution of electrostatic potential in the ionosphere that best fits all the line-of-sight velocity measurements. Figure 3.18 shows an example of such a reconstruction for the northern hemisphere. These convection maps can be generated at one or two minute resolutions of the radar scans illustrating the potential of the use of SuperDARN electric potential patters for improving the agreement between the various existing ionospheric models and actual observations.

## 3.5 In-Situ Measurements and Trans-Ionospheric Propagation

In-situ measurements and ionospheric sounding by Global Navigation Satellite System (GNSS) signals have provided vital support to ground-based observations of the Earth’s upper atmosphere, ionosphere, and beyond into open space. Their common denominator is the use of the sophisticated instruments on board satellites for research that covers a diversity of solar-terrestrial conditions at all altitudes, latitudes, and longitudes.

### *3.5.1 Topside Measurements*

Observations of ionospheric parameters at altitudes above the height of maximum electron density $N(h)$ in the ionospheric F region (from about 300 to 1,000 km), often called the topside ionosphere, have helped in key discoveries of its prominent features that were not easy to achieve from ground-based experiments. Only after 1960 did it become possible to study the topside ionosphere using instruments on board rockets and satellites. Today, these in-situ measurements of space plasma from

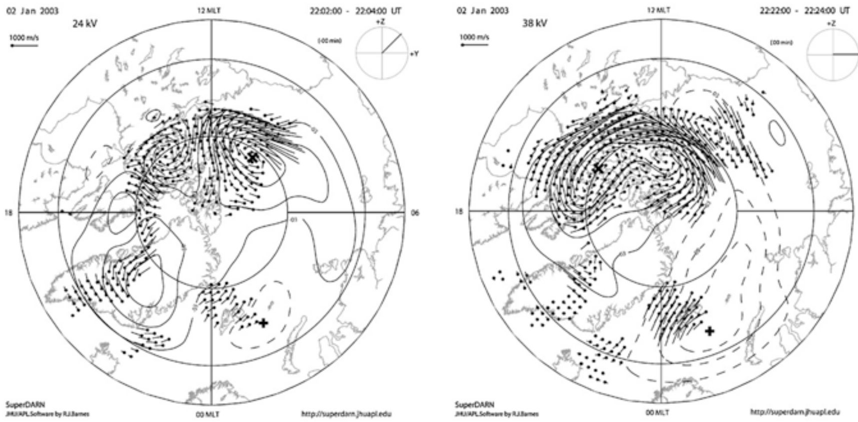


**Fig. 3.18** An example of SuperDARN MLT-MLAT polar *plots* showing northern hemisphere convection patterns for northward (*left panel*) and southward (*right panel*) IMF on 2 January 2003 (from Alfonsi et al. 2008)

various space platforms have advanced to the extent that it is possible to measure a number of important physical parameters in the terrestrial thermosphere, ionosphere, plasmasphere, magnetosphere, and interplanetary medium, as well as the upper atmosphere of other planets. Here only two main techniques are briefly described from among the many available, these being the earliest ones performed on sounding rockets and satellites and very frequently implemented.

The first technique makes use of a Langmuir probe and is a good example of a direct method for in-situ measurements. The second, utilizing topside ionosonde sounding equipment on a satellite, is an excellent example of an indirect method for in-situ measurements. Considering their fundamental role for establishing global ionospheric morphology and its variations with latitude, local time, seasons, and solar activity, they are highly relevant to the main objectives of this book. This is particularly true for the important role that data from satellite topside sounders have played in the development of empirical ionospheric models. Here the main emphasis is on the International Reference Ionosphere (IRI) model, for which these in-situ measurements are currently being used to improve the topside part of the IRI real-time mode.

Langmuir probes are devices carried on space vehicles and designed to measure plasma number density and temperature. Basically, they are conductors with an applied voltage that are immersed in space plasma, from which a current is drawn. It is the measurement of this current that provides an estimation of the plasma number density and temperature. There are different types of Langmuir probes depending on applications, including those dedicated to electron temperature measurements. In principle, they are either swept in voltage (Swept Langmuir Probe) or operate at a fixed DC voltage (Fixed-Bias Langmuir Probe). However, a serious difficulty in all measurements with Langmuir probes is that the space plasma being studied is disturbed by the vehicle moving through it. This problem is not relevant to ionospheric prediction and forecasting or radio propagation, but it is extremely important in plasma physics and is studied in a number of papers on that subject [see for example a review paper by Pfaff (1996)].

Perhaps the best known results are Langmuir probe measurements of electron plasma density *Ne* and electron temperature *Te* from a swept probe on the polar-orbiting Dynamics-Explorer-2 satellite at both high- and low-latitudes, showing large variations in *Ne* and *Te* on a regional scale in the Earth's ionosphere.

The topside sounding method is based on a radio wave propagation technique using an ionosonde, similar to the use of an ionosonde on the ground as described in Sect. 3.3, with an in-situ transmitter or an in-situ receiver on a satellite above the height of maximum electron density *N(h)* in the F region. The topside ionograms are recorded when the transmitted radio pulses travel downwards and return to the satellite after reflection. As they split into ordinary and extraordinary components and suffer lateral deviation, it is the vertical component of the group velocity that determines the time of travel. A topside ionosonde transmits signals sweeping through a typical frequency range from 0.1 to 10 or 20 MHz, and records the time delay of the ionospheric echoes. The topside ionogram established in this way usually includes an ordinary and an extraordinary reflection trace and its analysis requires finding these traces and then inverting them into an electron density profile. While the transmitter and receiver are enveloped in ionospheric plasma, it is assumed that the ionosphere is horizontally stratified so that the wave normals are vertical everywhere and that the electron density *N*(*h*) increases monotonically as the height *h* decreases, reaching the maximum of the F region (Fig. 3.19).

Although current topside ionograms have shown many new features covering large geographical areas sounded at relatively short time intervals with signals which do not suffer D region absorption, they are never simple to interpret

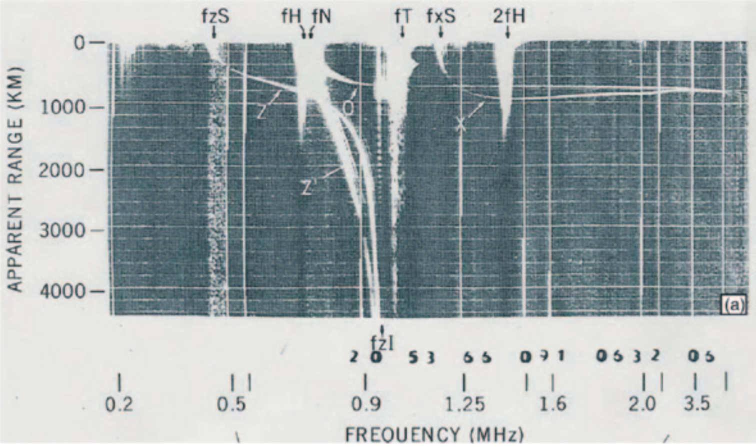


**Fig. 3.19** An Alouette 2 mid-latitude ionogram, in negative format with signal reception in *white* on a *black* background showing Z, O, and X traces (from Benson 2010)

(Fig. 3.19). There are many reasons for this even during quiet ionospheric conditions, such as a non monotonic electron density $N(h)$ distribution and the fact that when the ionosphere is not horizontally stratified there would be obvious oblique reflections. When the ionosphere is disturbed the real situation is even more complicated with ionogram traces being broadened and irregular. Sometimes topside ionograms show resonances or spikes at certain frequencies as vertical lines of such intensity that the true parameters are difficult to read.

Probably the most successful sounders have been the Canadian-built and US-launched, Alouette 1 and 2, then the ISIS 1 and 2 and the Japanese ISS-b. The Ionosphere Sounding Satellite-b (ISS-b) launched in February 1978 conducted topside sounding during the solar maximum period from 1978 through 1980 and produced significant results regarding the global distribution of ionospheric parameters like the manually scaled critical frequency *foF2*. The global maps of *foF2* obtained from this topside sounding experiment were compared with very-well known *foF2* maps generated by the CCIR and URSI models making important contributions to long-term ionospheric prediction. It was shown that the new URSI model represents the ISS-b maps better than the older CCIR model, which will be discussed further in Chap. 6.

### 3.5.2 Ionospheric Sounding with GNSS Signals

The ionospheric total electron content (usually abbreviated to *TEC*) provides an overall description of ionization in the ionosphere-plasmasphere system. Therefore, it is one of the most important ionospheric parameters not only for trans-ionospheric communications (propagation at VHF and above), navigation, and solar-terrestrial physics, but also for a number of modern technical and technological applications. The ionospheric total electron content is defined as the integral of electron density along the radio wave path $s$ from a satellite transmitter to a ground-based receiver:

$$TEC = \int_s Ne\,(s)\,ds \tag{3.5.1}$$

where $Ne$ is the ionospheric electron density along the path $s$ in electrons/m$^3$. As the $s$-axis stands for the receiver-to-satellite direction, the $ds$ is the element of its path. This so-called slant *TEC* can be converted to the equivalent vertical total electron content *VTEC* at its associated longitude $\lambda$ and latitude $\varphi$, *TEC* along a trans-ionospheric ray path vertically above the location of the receiver, by using different model assumptions about the horizontal structure of the ionosphere. Following Fig. 3.20, it is clear that *TEC* values for a vertical path can be obtained from total slant *TEC* using the relationship:

$$Ne\ (s)\,ds = Ne\,(h)\ \sec\chi\, dh \tag{3.5.2}$$

where $\chi$ is the zenith angle of the ray at the mean ionospheric height (here estimated at 400 km).

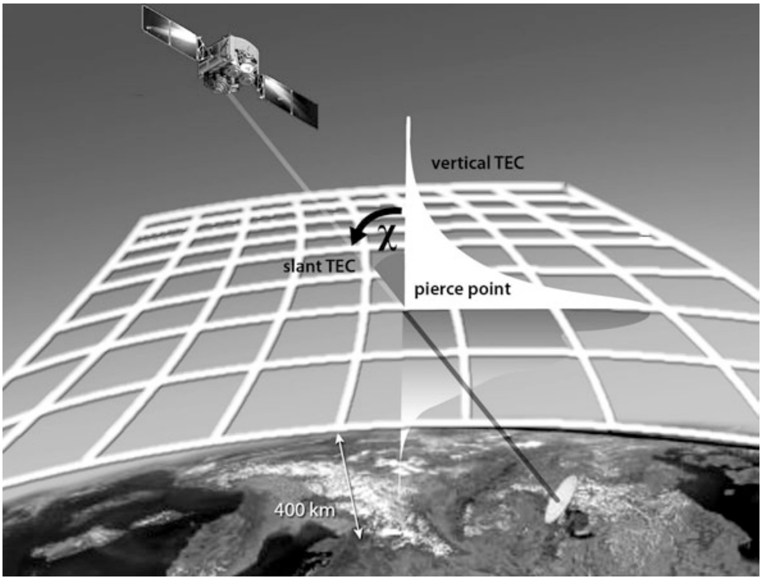


**Fig. 3.20** Geometry of *TEC* for a vertical path in relation to the total slant *TEC*

The *TEC* represents the number of free electrons in a column of unit cross-section taken along the propagation path from the receiver to the satellite, which can be an oblique or slant path (*STEC*) or a vertical path (*VTEC*). For convenience, the unit used for both *TECs* is the quantity of $10^{16}$ electrons/m$^2$, called a total electron content unit or TECU. Combining ionosonde data with overhead *TEC* data, the ratio between the vertical total electron content and the electron density at the peak of the F2 layer (*Nmax*) is known as the ionospheric slab thickness:

$$\tau = VTEC/Nmax = VTEC/1.24\,10^{10}\,(foF2)^2 \tag{3.5.3}$$

where *foF2* is given in MHz, *VTEC* in electrons/m$^2$ and $\tau$ in meters. This represents the thickness of a hypothetical ionosphere of uniform electron density *Nmax* and electron content *TEC* reflecting the shape of the ionospheric height profile.

Since the change in satellite-to-receiver signal propagation time due to the anisotropic nature of the ionosphere-plasmasphere system is directly proportional to the total number of free electrons along the signal path, it is obvious that *TEC* could have been measured along radio links from a satellite transmitter to a ground receiver since the 1960s by a range of different techniques. Some early *TEC* measurements using low-orbiting or geostationary satellites utilized the ionospheric Faraday effect as a technique for determining the polarization rotation of VHF signals. Subsequently, a declining number of Faraday satellites and the rapid implementation of various satellite navigation and positioning systems, made it possible to use the ground-based NNSS (Navy Navigation Satellite System) measurement technique, and then since 1994 the Global Positioning System (GPS).

More recently, the GNSS (Global Navigation Satellite System) network of receivers tracking the transmissions of modernized US GPS, the Russian GLONASS, and the planned European Galileo and Chinese COMPASS systems have revolutionized ionospheric science by providing a new source of ionospheric data that is globally distributed and continuously available. The highly accurate GPS data in the RINEX (Receiver INdependent EXchange) format are available at several global and regional centers of the International GNSS Service formerly the International GPS Service for Geodynamics (IGS) (http://igscb.jpl.nasa.gov/). These data allow regular *TEC* determination along satellite-receiver links over certain areas with a broad range of time resolution staring from 30 s by different processing techniques. One of the techniques for *TEC* evaluation from RINEX files is described in Chap. 7, as these data have been used to produce global, regional, and/or local maps of ionospheric total electron content in retrospective and real-time modes.

The ionosphere propagation delay is the largest error source for the single GNSS frequency, and the second for differential GPS (http://www.trimble.com/gps/howgps-error2.shtml#3). The amount of delay affecting a signal in the band of frequencies used by the GNSS is inversely proportional to the square of its carrier frequency but proportional to *TEC* along the ray path. Therefore, for the electromagnetic waves transmitted by GPS satellites on two L-band frequencies ($L1 = 1{,}575.42$ MHz and $L2 = 1{,}227.60$ MHz), the ionospheric phase delay (also called the pseudo-range error) expressed as a distance corresponding to the apparent increase in path length, has been estimated in a first-order approximation as:

$$d_I = KTEC/f^2 \tag{3.5.4}$$

where $d_I$ is in meters, $K$ is a constant equal to 40.3 $\mathrm{m}^3\,\mathrm{s}^{-2}$ and $f$ is the carrier frequency of the signal in Hz.

The range error caused by 1 TECU at *L1* frequency is about 0.162 m. With tens of *TEC* units along a GNSS signal path through the ionosphere during different solar-terrestrial conditions, ionosphere delays can account for position errors in the order of tens of meters. The well known approach to reduce the positioning error of single frequency GNSS receivers is to use ionospheric models. However, ionospheric models vary in complexity, and even state-of-the-art models can correct, in general, for only 70 to 80 % of the delay because of the high ionospheric variability. It will be demonstrated in Chap. 7 that the best correction for the effects of ionospheric delay can be achieved using an actual measurement. Incorporating a near real-time ionospheric monitoring capability at an observation location can eliminate 90–95 % of the ionospheric time delay error.

## Suggested Readings

Alfonsi L, Kavanagh AJ, Amata E, Cilliers P, Correia E, Freeman M, Kauristie K, Liu R, Luntama J-P, Mitchell CN, Zherebtsov GA (2008) Probing the high-latitude ionosphere from ground-based observations: The state of current knowledge and capabilities during IPY (2007–2009). J Atmos Sol-Terr Phys 70:2293–2308

Appleton EV, Builder G (1933) The ionosphere as a doubly refracting medium. Proc Phys Soc 45:208–220

Benson RF (2010) Four decades of space-borne radio soundings. Radio Sci Bull 333:24–44

Bowles KL (1958) Observations at vertical incidence scatter from the ionosphere at 41 Mc/sec. Phys Rev Lett 1:454–455

Budden KG (1961) Radio waves in the Ionosphere. Cambridge University Press, Cambridge

Chisham G, Lester M, Milan SE, Freeman MP, Bristow WA, Grocott A, MacWilliams KA, Ruohoniemi JM, Yeoman TK., Dyson P, Greenwald RA, Kikuchi T, Pinnock M, Rash J, Sato N, Sofko G, Villain J-P, Walker ADM (2007) A decade of the Super Dual Auroral Radar Network (SuperDARN): scientific achievements, new techniques and future directions. Surv Geophys 28:33–109

Ciraolo L, Spalla P (1997) Comparison of total electron content from Navy Navigation Satellite System and the GPS. Radio Sci 32:1071–1080

Davies K (1990) Ionospheric Radio. IEE Electromagnetic Waves Series 31, Peter Peregrinus Ltd, London

Dominici P (1971) Radio Propagazione Ionospherica. Supplement to Vol. XXIV of Ann Geofis

Evans JV (1969) Theory and practice of ionosphere study by Thomson scatter radar. Proc IEEE 57:496–530

Galkin IA, Reinisch BW (2008) The new ARTIST 5 for all digisondes. Iono Net Ad Group Bull 69:8

Gordon WE (1958) Incoherent scatter of radio waves by free electrons with applications to space exploration by radar. Proc IRE 46:1824–1829

Hunsucker RD (1991) Radio techniques for probing the terrestrial ionosphere. Springer, New York

Lilensten J, Cander LjR (2003) Calibration of the TEC derived from GPS measurements and from ionospheric models using the EISCAT radar. J Atmos Sol-Terr Phys 65:833–842

Kaplan ED (1996) Understanding GPS Principles and Applications. Artech House, Boston

Klobuchar J (1997) Real-time ionospheric science: the new reality. Radio Sci 32:1943–1952

McNamara LF (1991) The Ionosphere: Communications, Surveillance, and Direction Finding. Krieger Publishing Company, Malabar

Pezzopane M, Scotto C (2007) The automatic scaling of critical frequency foF2 and MUF(3000) F2: a comparison between Autoscala and ARTIST 4.5 on Rome data. Radio Sci. doi:10.1029/2006RS003581

Pezzopane M, Scotto C (2008) A method for automatic scaling of F1 critical frequencies from ionograms. Radio Sci. doi:10.1029/2007RS003723

Pfaff RF (1996) In-situ measurement techniques for ionospheric research. In: Kohl H, Ruster R, Schlegel L (eds) Modern Ionospheric Science. European Geophysical Society, Kaltenburg-Lindau

Piggot WR, Rawer K (1972) U.R.S.I Handbook of Ionogram Interpretation and Reduction. World Data Center A for Solar-Terrestrial Physics-Report UAG-23, NOAA, Environmental Data Service, Asheville

Ractliffe JA (1962) The Magneto-Ionic Theory and its Applications to the Ionosphere. Cambridge University Press, Cambridge

Rawer K (1956) The Ionosphere. Frederick Ungar Publishing Co, New York

Rawer K (1993) Wave Propagation in the Ionosphere. Kluwer Academic Publishers, Dordrecht

Reinisch BW, Huang X (1983) Automatic calculation of electron density profiles from digital ionograms 3. Processing of bottom-side ionograms. Radio Sci 18:477–492

Reinisch BW, Galkin IA, Khmyrov GM, Kozlov AV, Bibl K, Lisysyan IA, Cheney GP, Huang X, Kitrosser DF, Paznukhov VV, Luo Y, Jones W, Stelmash S, Hamel R, Grochmal J (2009) The New Digisonde for research and monitoring applications. Radio Sci. doi: 10.1029/2008RS004115

Stankov SM, Stegen K, Muhtarov P, Warnant R (2011) Local ionospheric electron density profile reconstruction in real-time from simultaneous ground-based GNSS and ionosonde measurements. Adv Space Res 47:1172–1180

## Additional Web Site

http://www.ips.gov.au/IPSHosted/INAG/web-69/2008/artist5-inag.pdf

# Chapter 4
# Ionospheric Spatial and Temporal Variations

## 4.1 Introduction

It was obvious from the time of its discovery that the Earth's ionosphere is highly variable in space and time. Its morphology affects propagation via the ionosphere and is characterized by geographical and geomagnetic variations on one hand, and daily, seasonal and solar cycle variations on the other.

The nature of ionospheric effects is not a simple function of the location of a receiver, but of the geographic and geomagnetic extent of the regions crossed by the transmission path. For a HF radio link with a single reflection the ionospheric properties at the mid-point of the path are important. In the case of a GNSS path it is the sub-ionospheric point, and so receivers located in the lower mid-latitudes can be affected by the ionosphere in the equatorial region, just as receivers in the higher mid-latitude regions can be affected by the ionosphere of the auroral region. Consequently, the ionosphere of the low and high-latitude regions can affect propagation beyond the boundaries of these regions.

The following discussion first addresses the main geographic regions of the ionosphere, and then focuses on daily, seasonal and solar cycle variations in the main characteristics and parameters within these regions. Emphasis is given to the mid-latitudes, since the processes observed in this geographical region of the ionosphere can be applied in general over a wider range of latitudes, and for the technologically most important part of the Earth's ionosphere, the F region.

The numerous characteristics and parameters obtained from measurements described in Chap. 3 define the main features of the ionospheric layers and their geographical and geomagnetic variations (Sect. 4.2). As explained in Chap. 2, only the E, F1, sporadic E when present, and F2 layers refract HF waves and can be subject to prediction, real-time specification or nowcasting, and forecasting for propagation purposes. The D region does not refract HF radio waves but plays a central role in their absorption, and so it is partly considered in Sect. 4.3.1. The ionospheric characteristics of the E and F1 layers are considered in Sect. 4.3.2 in relation to empirical modelling, because they follow a simple Chapman model behaviour.

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_4,

With its persistence 24 h a day and high altitude role, allowing the longest communication paths and normal refraction of the highest frequencies in the HF range, the F2 layer plays a vital role in ionospheric and trans-ionospheric propagation. The main characteristics of this layer, with the longest electron lifespan and the associated vertical total electron content (*VTEC*) subject to significant daily, seasonal, and solar cycle variations, are described in detail for the mid-latitude ionosphere in Sect. 4.3.3. The solar cycle variations at mid-latitude relevant for this book are presented for all layers in Sect. 4.4.

## 4.2 Geographic and Geomagnetic Variations

The geographic regions of ionospheric behaviour can be divided into: (1) high-latitudes where geomagnetic latitude is 90° > λ > 60° on each side of the magnetic equator; (2) mid-latitudes where geomagnetic latitude is 60° > λ > 20° on each side of the magnetic equator; and (3) low-latitudes where geomagnetic latitude is 0° < λ < 20° on each side of the magnetic equator. Here it is important to recall that the geomagnetic and geographic equators do not usually coincide, and may be up to 12° apart. At Asian longitudes the geomagnetic equator is the above the geographic equator, whereas at American longitudes it is below.

Some of these main regions are further divided into sub-regions such as the auroral and polar cap regions in high-latitudes and the equatorial and equatorial "anomaly" regions in low-latitudes. Figure 4.1 shows the approximate geographic extent of each of these main regions during quiet geomagnetic conditions when the mid-latitude ionosphere incorporates the transitional regions.

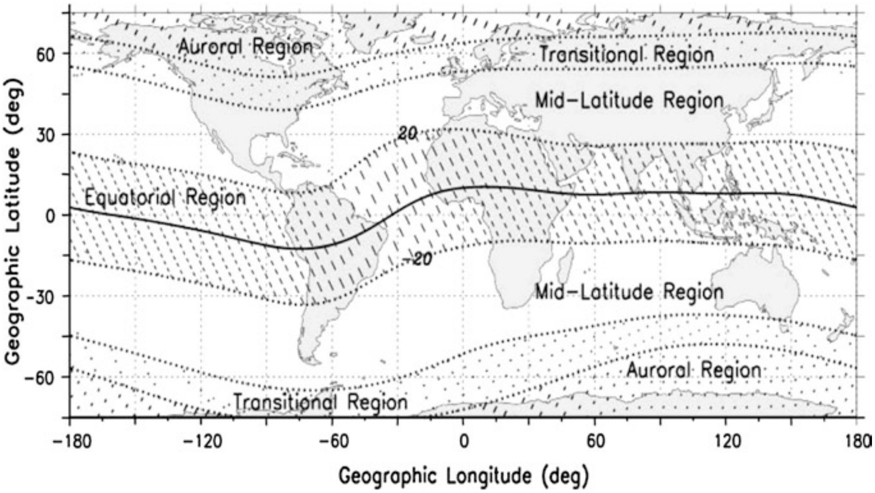


**Fig. 4.1** The major geographic regions of the ionosphere

During disturbed geomagnetic conditions the auroral region expands towards the equator, reducing the size of the mid-latitude region. Usually auroral activity is regarded as a good indicator of increased connections between the Earth's magnetosphere, ionosphere, and atmosphere. However, a direct contact between the magnetosphere and the interplanetary magnetic field produces a cusp or cleft at geomagnetic latitudes from 78° to 80° around local noon time. This phenomenon of a few degrees in width is characterized by increased electron densities at all altitudes.

Overall, the ionosphere at mid-latitudes is less complex than in other regions under prevailing solar-terrestrial conditions and is characterized by relatively small and slowly varying spatial gradients. These prevailing conditions exist when the ionosphere is not disturbed and are referred to as the "normal" or "quiet" ionosphere, which is the case approximately 98 % of the time. During the remaining approximately 2 % of the time, geomagnetic storms cause the ionosphere to be disturbed (Fig. 4.2).

Consequently, the mid-latitudes are a relatively benign part of the Earth's ionosphere, free of the complications that occur at high-latitudes due to particle precipitation, and at low-latitudes where marked ionization irregularities and associated gradients are encountered. It is the best studied ionospheric region with the densest network of vertical sounding stations (at least in the northern hemisphere) dating back over more than 50 years. The most advanced technological societies are located in this area and it is reasonable to use this as a basis when discussing the behaviour of the ionosphere for prediction and forecasting purposes. Therefore, in Sect. 4.3 regular daily, seasonal, and solar cycle variations of the main ionospheric characteristics and parameters at mid-latitudes are presented.

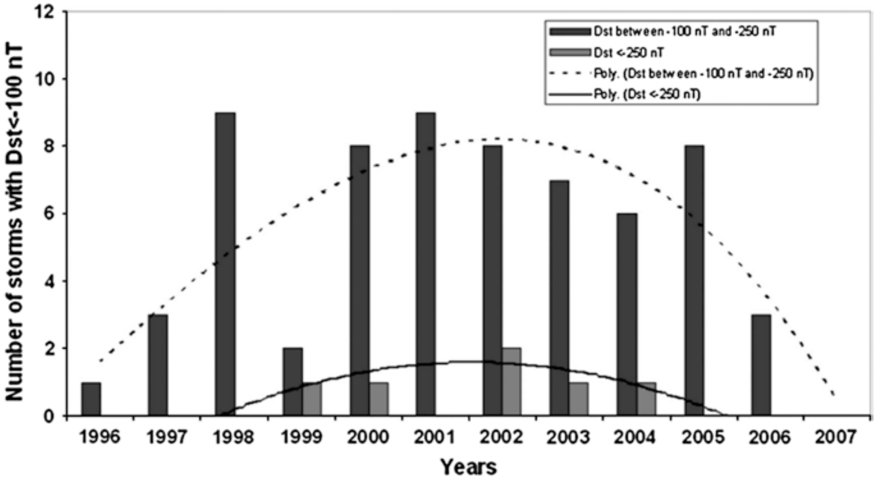


**Fig. 4.2** Yearly count of geomagnetic storms during solar cycle 23 (1996–2007)

### *4.2.1 High-Latitude Ionosphere*

The high-latitude ionosphere was subjected to intensive studies before and during the International Geophysical Year (IGY 1957–1958) and more recently during the International Polar Year (IPY 2007–2009). These studies confirmed that the high-latitude ionosphere is, understandably, subject to the highest variability of all the regions as a result of its connection, via the Earth's magnetic field, to the magnetosphere and the interplanetary medium (see Chap. 2).

Ion production by solar EUV radiation and X-rays is relatively weak because of the low elevation of the Sun, and energetic charged particles are the main contributors to ion production at high-latitudes. The polar ionosphere at geomagnetic latitudes greater than about 75° is also the region where most of the solar wind energy is dissipated and where this dissipation often dominates the total upper atmospheric energy budget. The energy is redistributed by heating and cooling, winds and waves.

These processes are extremely dynamic in nature and their global influence maximizes during geomagnetic storms and substorms. Hence ionospheric behaviour at high-latitudes is different from high-latitudes and predominantly irregular in all three regions D, E, and F. It makes studies aiming to establish a high-latitude climatological pattern extremely difficult and favours identification and interpretation of the morphological features during diverse disturbed conditions, as will be described in Chap. 8.

An important feature of high-latitudes is the main ionospheric trough (MIT) found in the electron density of the night-time F region. It is the natural boundary between the mid-latitude and auroral ionosphere, which initially exists in the afternoon sector around 70° magnetic latitudes as ionized patches of a few degrees in width. The MIT shifts slowly to about 57° around local midnight (during moderate geomagnetic activity) reaching around 10° or more in width as local time increases. The ratio of densities inside and outside the trough can vary between less than 0.1 to almost 1. The latitudinal gradients are much more pronounced at the topside ionosphere.

In Fig. 4.3 some trough examples for the satellite Dynamic Explorer DE 2 are shown. It is evident that trough modelling is important for ionospheric real-time specification at mid-latitudes, although its forecasting can be rather difficult. For long-term prediction, representing monthly median conditions of electron density, such phenomena can be ignored.

### *4.2.2 Equatorial Ionosphere*

The equatorial ionosphere, particularly its F region, is subject to a wider range of peculiar behaviour when compared to other latitude regions. This is mainly because the Sun is more directly overhead and the Earth's magnetic field is

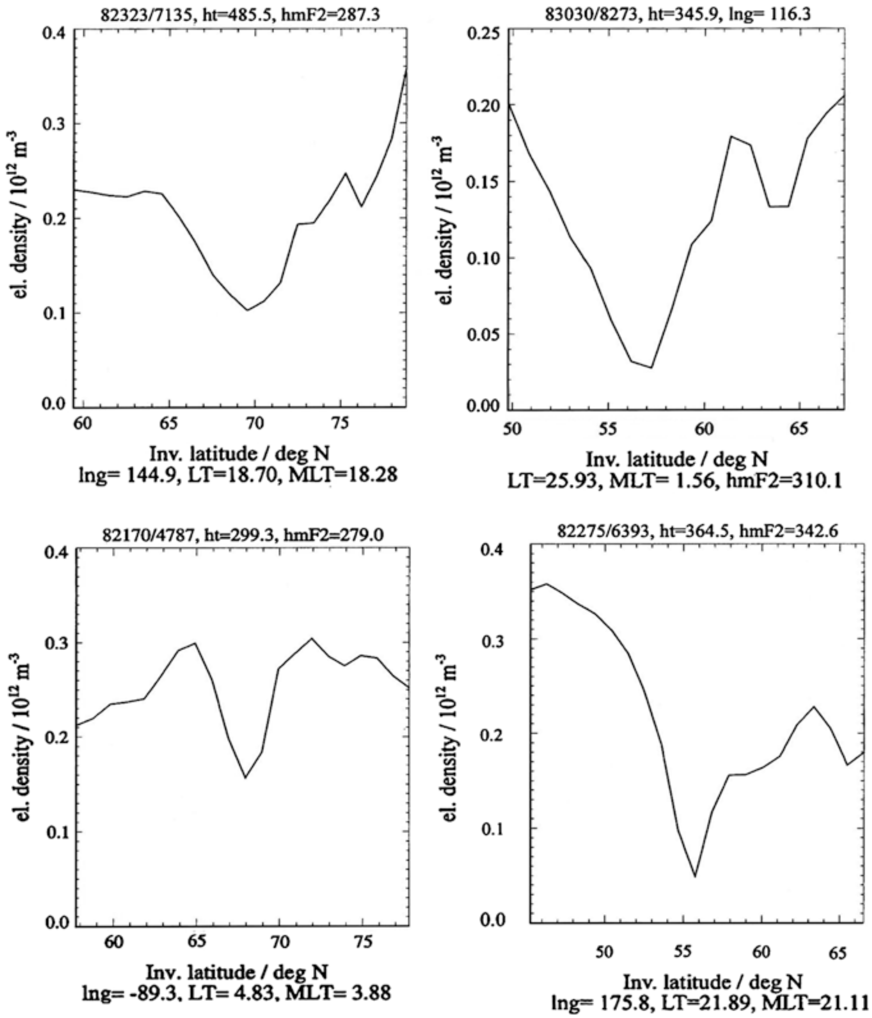


**Fig. 4.3** Examples for DE 2 troughs projected to the F2 peak for winter conditions (*top panels*), summer conditions (*bottom left panel*) and equinox conditions (*bottom right panel*). DE 2 orbit parameters are found in the headers and foot lines of each graphic (from Feichter and Leitinger 2002)

horizontal at the geomagnetic equator thus electric fields become more influential causing significant vertical drifts of ionization. Therefore, the equatorial ionosphere is higher in altitude than at mid-latitudes, with the electron density falling to a minimum near the magnetic equator and passing through maxima at latitudes ±15° to ±20° on each side of the magnetic equator. This phenomenon is called the Appleton anomaly or just the Equatorial Ionization Anomaly (EIA). It is contrary to the simple ionospheric model, which assumes electron density to be maximum over the equator because the Sun is overhead with the maximum solar ultraviolet (UV) ionization rate.

In general the ionospheric plasma motion is a combination of electro dynamic drift and thermal diffusion, with the net result often described as a fountain effect as free electrons migrate upwards in altitude and then away from the magnetic equator towards higher latitudes. Figure 4.4 shows the latitudinal variations of the F2 layer maximum electron density *NmF2* and the corresponding variations of height *hmF2* obtained from ISIS-1 ionograms recorded on 27 October 1969 at 160° E at 1505 LMT. The timing of the anomaly development, the magnitude in terms of the ratio of the mean *foF2* at two maxima, north and south of the magnetic equator relative to the *foF2* at the magnetic equator, and its dependence on geomagnetic activity, vary significantly.

As already mentioned in Sect. 2.5.2, equatorial irregularities in the F region are considerable, causing strong amplitude scintillation fading and phase scintillation effects on L-band satellite to ground trans-ionospheric signals, but allowing radio wave trans-equatorial propagations far beyond those usually seen in the HF range.

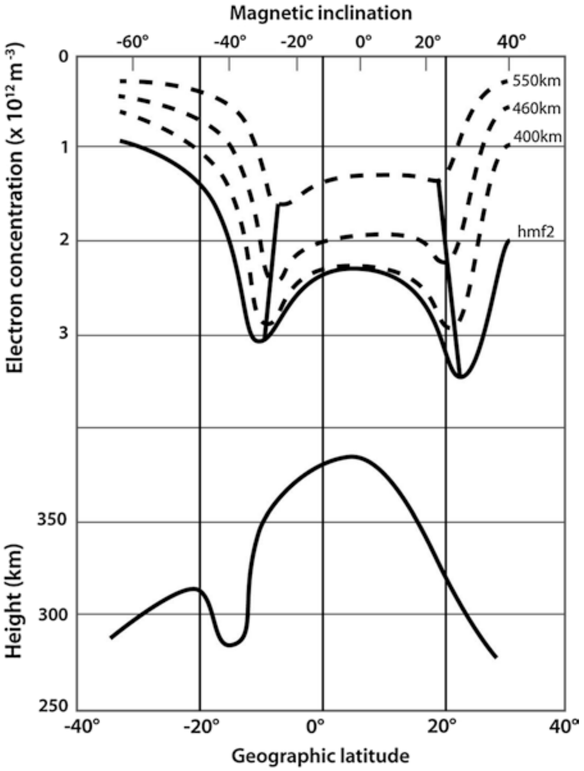


**Fig. 4.4** Latitudinal variations of *NmF2* (*continuous curve*) and electron concentration at fixed heights in the topside ionosphere (*broken curves*) and corresponding variation of *hmF2*

Usually during post-sunset local time, while changes in the height of the F region above the dip equator are underway, ionization irregularities of about 3 m to more than 10 km in size tend to develop, mostly disappearing by morning. They appear in the form of expanding plumes or different scale bubbles of local ambient and can be seen on ionograms as a thickening or spreading of the F region trace (see Fig. 2.26).

Here it is important to note that the occurrence of irregularities appears to decrease during geomagnetically active periods when $Kp > 4$. The similarity with the equatorial anomaly, well-developed during geomagnetically quiet days and absent when $Kp > 5$, suggests a need for profound investigation of the complex physical processes operating at equatorial latitudes that is beyond the scope of this book.

## 4.3 Daily and Seasonal Variations at Mid-Latitudes

First, it is very important to note that historically, ionospheric theory is based on the Chapman equations for the production of an ionospheric layer and their description of electron densities under conditions of photochemical equilibrium. On the basis of this theory, seasonal, local time, and latitudinal variations of electron density are the Chapman functions of solar zenith angle (see Chap. 2). Although this is valid for the lower ionospheric layers such as D, E, and F1, the peak electron density of the F2 layer, *foF2*, behaves very differently from the simple Chapman theory. Accordingly, any behaviour of the ionospheric F2 layer not consistent with solar zenith angle dependence was considered anomalous. Nowadays, such behaviour is attributed to physical processes, as transport of ionospheric ions and electrons, variations in atmospheric composition, and other factors not included in the Chapman equations. The best examples of these so-called anomalies described in this Chapter are the winter (seasonal) and equatorial anomalies.

### 4.3.1 D Region

Although the ionospheric D region exhibits marked solar control, the form of its diurnal behaviour depends on the measuring technique used for observations. For instance, when VLF (very low frequency 3–30 kHz) radio waves are reflected in the D region the apparent height of reflection ($h$) varies according to Chapman theory for the height variation of a level of constant electron density:

$$h = h_0 + H \ln \sec \chi \tag{4.3.1}$$

where $h_0$ is about 72 km and $H$ is the scale height of the neutral atmosphere of about 5 km. However, at oblique incidence the diurnal pattern is quite different, with a sharp reduction of reflection height before sunrise, and a fairly rapid recovery after sunset. Such behaviour is related to the complex photochemistry of the D region, which is beyond the scope of this Chapter and will be briefly considered in related to the absorption properties of the D region in Chap. 9.

### 4.3.2 E and F1 Layers

The E layer plasma frequency dependents on the solar zenith angle $\chi$ and solar activity expressed by the monthly median sunspot number $R$ following the empirical law:

$$foE(\chi, R) = 3.3\,[(1 + 0.0088R)\,\cos\chi]^{1/4}\ \text{MHz} \tag{4.3.2}$$

meaning that the peak electron density $NmE$ varies as $(\cos\chi)^{1/2}$.

On average the electron density in the E layer corresponds to plasma frequency of ~3 MHz during the daytime. The night-time E layer remains weakly ionized with corresponding plasma frequencies between 0.4 and 0.6 MHz, which makes classical ionosonde measurements impossible. As an example of typical diurnal and seasonal variations of the E layer, Fig. 4.5 shows the monthly median *foE* values for the summer month of July, and the winter month of December during high (2000) and low (2008) solar activity at the Chilton (51.6° N, 358.7° E) ionosonde station. The *foE* is higher in summer than in winter under any solar-terrestrial

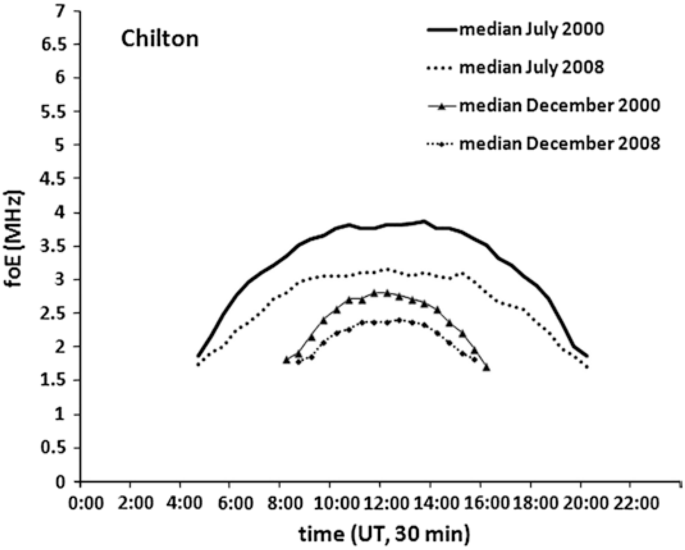


**Fig. 4.5** Diurnal variations of the *foE* monthly median values during July and December at high (2000) and low (2008) solar activity at the Chilton (51.6° N, 358.7° E) ionosonde station

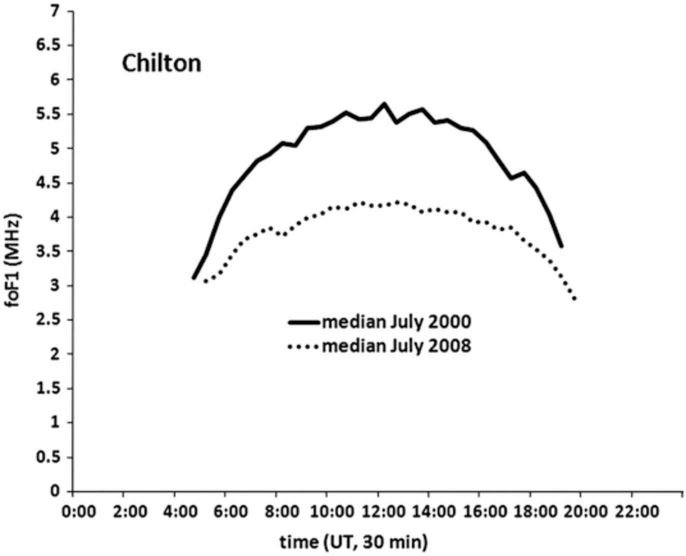


**Fig. 4.6** Diurnal variations of the *foF1* monthly median values during July at high (2000) and low (2008) solar activity at the Chilton (51.6° N, 358.7° E) ionosonde station

conditions. In addition, this difference is particularly pronounced under a similar solar activity level (see July/December in 2000 and July/December 2008, respectively in Fig. 4.5).

The *foF1* shows a dependence on solar zenith angle $\chi$ and the monthly median sunspot number $R$ similar to the E layer dependence:

$$foF1(\chi, R) = 4.25\,[(1 + 0.015R)\,\cos\chi]^{1/4}\ \mathrm{MHz} \tag{4.3.3}$$

indicating $\alpha$-Chapman behaviour.

On average the electron density in the F1 layer corresponds to a plasma frequency of ~4.5 MHz during the daytime. The F1 layer itself was never seen in winter at high solar activity in 2000, and only occasionally at low solar activity in 2008. Figure 4.6 shows the monthly median *foF1* values for the summer month of July during high (2000) and low (2008) solar activity at the Chilton ionosonde station. The *foF1* is significantly higher in July 2000 than in July 2008, as expected.

### *4.3.3 F2 Layer*

The characteristics of the ionospheric F2 layer vary much more significantly by solar cycle, season, and local time than any other layers, resulting not only from changes in solar EUV radiation and X-ray, but also from various chemical and electrodynamic processes. As such, these are essential for understanding monthly median ionospheric behaviour, often defined as ionospheric climatology, and its prediction as described in Chap. 6. At the same time they form a significant part of ionospheric storm behaviour, often defined as ionospheric weather, and its timely forecasting as described in Chap. 8. Of particular interest for this book are the *foF2, M(3000)F2*, and *VTEC* as previously defined.

#### 4.3.3.1 Daily and Seasonal Variations

Considered only according to simple Chapman theory (without plasma transport and assuming photochemical equilibrium), the diurnal and seasonal variations of the F2 layer appear anomalous in many ways. The major anomalies of the F2 layer at mid-latitudes, well described in literature, are as follows: (1) the December anomaly characterized by *foF2* and *NmF2* (proportional to square of *foF2*) values systematically higher all over the world in the ionospheric winter season (November, December, and January), than in the ionospheric summer season (May, June, and July); (2) the winter or seasonal anomaly characterized by daytime *foF2* and *NmF2* values in high- and mid- latitudes two and four times higher, respectively, in winter than in summer. These two anomalies are in phase on the northern hemisphere and separated by half a year on the southern hemisphere; (3) the semi-annual anomaly characterized by *foF2* and *NmF2* values unusually high at the ionospheric equinoxes (March, April, September, and October).

Ionospheric day-to-day variability by season presents an initial problem of scientific and practical importance to be considered here at a single but typical mid-latitude station: Chilton (51.6° N, 358.7° E), an ionosonde that has been operating for decades starting at the Slough (51.5° N, 359.4° E) site in the UK. It is important to note that the longest historical sequence of ionosonde measurements is from the Slough station, with data available from 1932 onwards.

Figure 4.7 shows the UT (here equal to LT) variations of the F2 layer critical frequency *foF2* in four ionospheric seasons at Chilton in 2000, which was the year of maximum solar activity in the 23rd solar cycle. Each ionospheric season is represented by one calendar month, namely March for the spring and October for the autumn equinox seasons, July for the summer and December for the winter seasons. With observed monthly mean values of *Ri* in the range of 138.5 in March, 170.1 in July, 99.4 in October, and 104.4 in December 2000 these data represent high solar activity. With the observed monthly mean values of the *Ap* index in the range of 9 in March, 21 in July, 18 in October, and 7 in December 2000 these data represent low to medium geomagnetic activity, apart from July and October which were considerably disturbed months.

The principal systematic features emerging from Fig. 4.7 are: (1) a large summer/winter difference well defined in *foF2* values; (2) a strong semi-annual pattern characterized by the equinox months March and October having higher median levels and broader distributions of day-to-day fluctuations than the other months; (3) a wide scattering at any time representing a considerable day-to-day

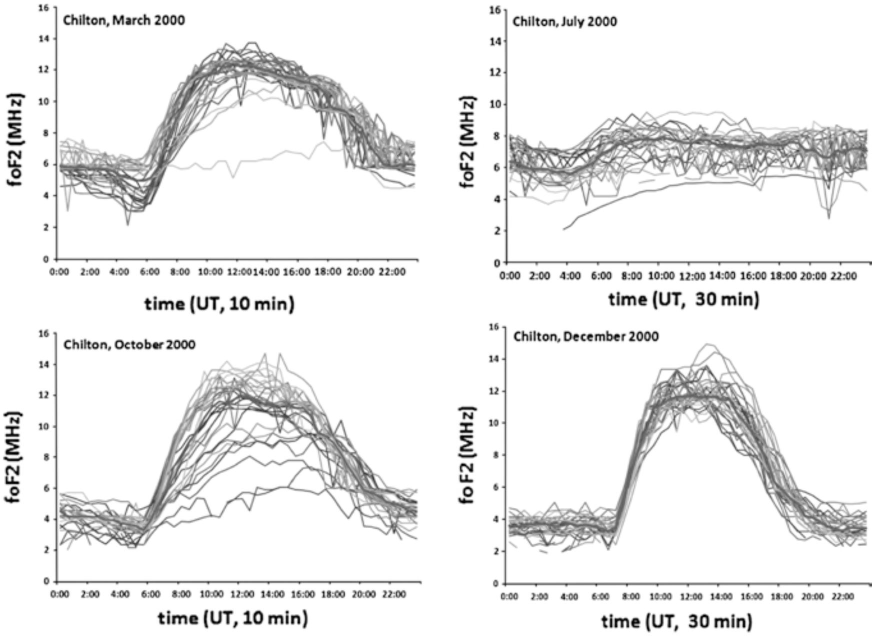


**Fig. 4.7** Variation of *foF2* at the Chilton (51.6° N, 358.7° E) ionosonde station for every day during March, July, October, and December 2000 (*bold lines* represent respective monthly median values)

variability of monthly median values, higher during daytime than at night. In all these plots the diurnal variation has a sharp, short-lived, and seasonally dependent early morning minimum, and a relatively slow decrease in *foF2* values approaching their night-time values.

The monthly median *foF2* values for these periods are shown in Fig. 4.8. The *foF2* is higher in winter than in summer, around 50 % from 08 to 17 UT, while *foF2* is opposite during night-time. In addition, the highest *foF2* values are observed during the equinoxes, with the spring *foF2* values (March 2000) especially pronounced. The difference is mostly due to the transitions between summer and winter.

For low solar activity, the diurnal and seasonal *foF2* variations differ considerably in some aspects. Figure 4.9 shows the UT variations of *foF2* in four ionospheric seasons at Chilton in 2008, which was the year of minimum solar activity. With observed monthly mean values of *Ri* in the range of 9.3 in March, 0.8 in July, 2.9 in October, and 0.8 in December 2008 these data represent very low solar activity. With observed monthly mean values of the *Ap* index in the range of 11 in March, 5 in July, 7 in October, and 4 in December 2008 these data represent low geomagnetic activity, except for March, which had a few disturbed days. The monthly median *foF2* values for these months are shown in Fig. 4.10.

Both Figs. 4.7 and 4.9 show that the notable scattering around the monthly median values and the winter anomaly are still present at solar minimum, with

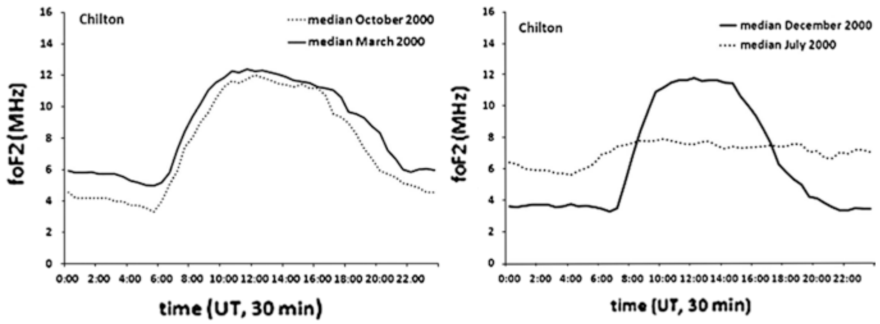


**Fig. 4.8** Diurnal variations of the *foF2* monthly median values at the Chilton (51.6° N, 358.7° E) ionosonde station during March and October 2000 (*left plot*), and July and December 2000 (*right plot*)

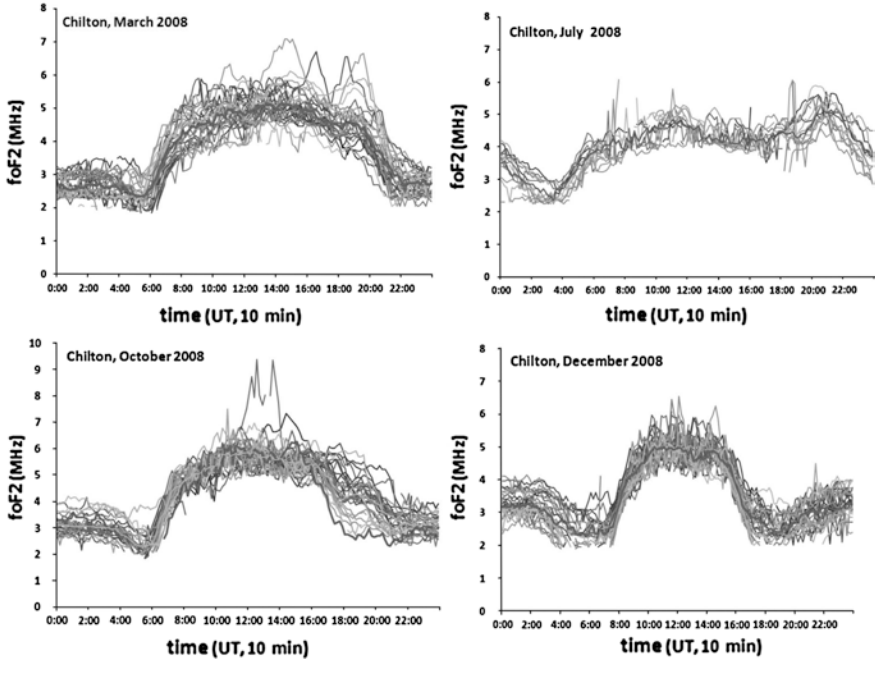


**Fig. 4.9** Variation of *foF2* at the Chilton (51.6° N, 358.7° E) ionosonde station for every day during March, July, October, and December 2008 (*bold lines* represent respective monthly median values)

daytime *foF2* values a little higher in winter than in summer. A new feature is the evening *foF2* peak in summer being clearly pronounced at low solar activity. Here the diurnal maxima of *foF2* are highest during the equinoxial months of March and October.

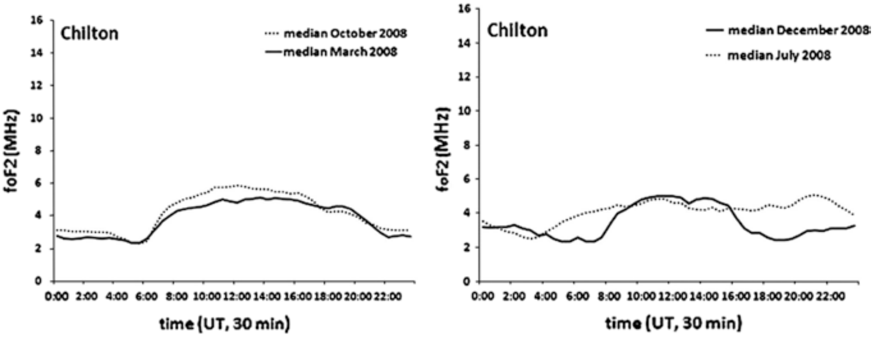


**Fig. 4.10** Diurnal variations of the *foF2* monthly median values at the Chilton (51.6° N, 358.7° E) ionosonde station during March and October 2008 (*left plot*), and July and December 2008 (*right plot*)

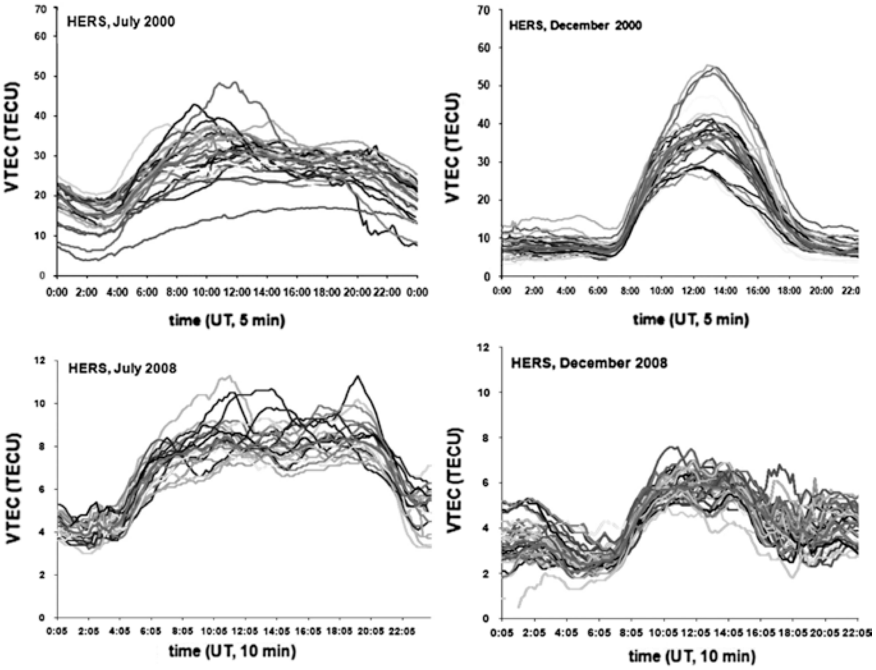


**Fig. 4.11** Plot of the *VTEC* diurnal variation over the mid-latitude GPS station HERS (50.9 N, 0.3 E) at solar maximum in July and December 2000 and solar minimum in July and December 2008 (*bold lines* represent respective monthly median values)

Total vertical electron content also varies diurnally, seasonally, geographically, and with the solar cycle. To examine the *VTEC* climatology at the GPS ground station at HERS (50.9 N, 0.3 E) in Hailsham, relatively close to the Chilton ionosonde station, the 5–10 min values are plotted in Fig. 4.11. The months of July and

December represent summer and winter conditions, respectively in high (2000) and low (2008) solar activity epochs. It can be seen from these plots that the *VTEC* vary significantly not only as a function of local time (here equal to UT) from day to day, but also by season and solar activity.

It is important to note that GPS measurements of *TEC* at a height of 20,000 km include electron content contributions from the D, E, F1 when present, and F2 layers of the ionosphere, and a quantity corresponding to the protonospheric *TEC* component. The protonosphere corresponds approximately to 10 % of the *TEC* of the combined ionosphere-protonosphere during daytime hours when electron density in the ionosphere is high, and 40–50 % of night-time *TEC*, when the ionospheric electron density is low.

Again as in case of *foF2*, at solar minimum in 2008 the daily *VTEC* differences within one month are significantly less than near solar maximum 2000. This is not surprising since daily *TEC* values are influenced not only by ionisation production but also the effect of geomagnetic activity that depends on season and magnitude. In addition, there is a fast and significant increase in *TEC* toward solar maximum activity. At mid-latitudes the night-time *VTEC* is about 5 TECU at low solar activity and the daytime *VTEC* is about 35 TECU at high solar activity as illustrated in Fig. 4.11. A well-developed seasonal anomaly during solar maximum is also seen in the left plot of Fig. 4.12. However, this is not the case during the solar minimum in 2008 (see right plot of Fig. 4.12).

As described in Sect. 3.3.4, the transmission *M* factor or obliquity factor *M(3000)F2* is one of the most important characteristics scaled from ionograms. The *foF2* and *M(3000)F2* product is equal to *MUF(3000)F2,* the Maximum Usable Frequency for a distance of 3,000 km for an F2 layer hop. It will be shown in Chap. 5 that the *M(3000)F2* values play a key role in several simple models of the height of the F region peak electron density, *hmF2*. Figure 4.13 shows the diurnal variation of *M(3000)F2* values at Chilton for every day during March, July, October, and December 2000, and it can be clearly seen that even at maximum solar activity almost all values of *M(3000)F2* fall within the narrow range of 2–4.

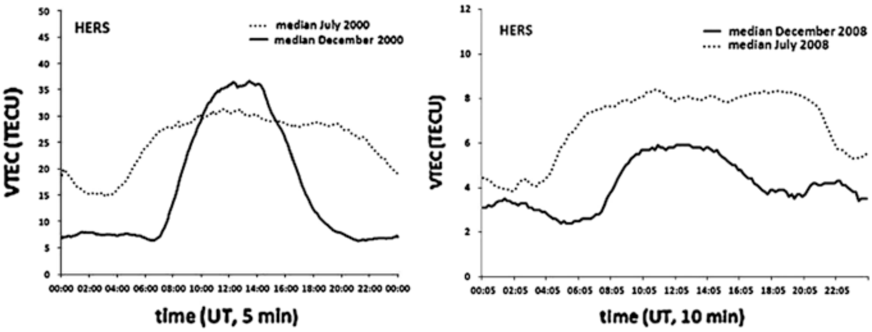


**Fig. 4.12** Diurnal variations of the *VTEC* monthly median values over HERS (50.9 N, 0.3 E) in July and December 2000 (*left plot*), and July and December 2008 (*right plot*)

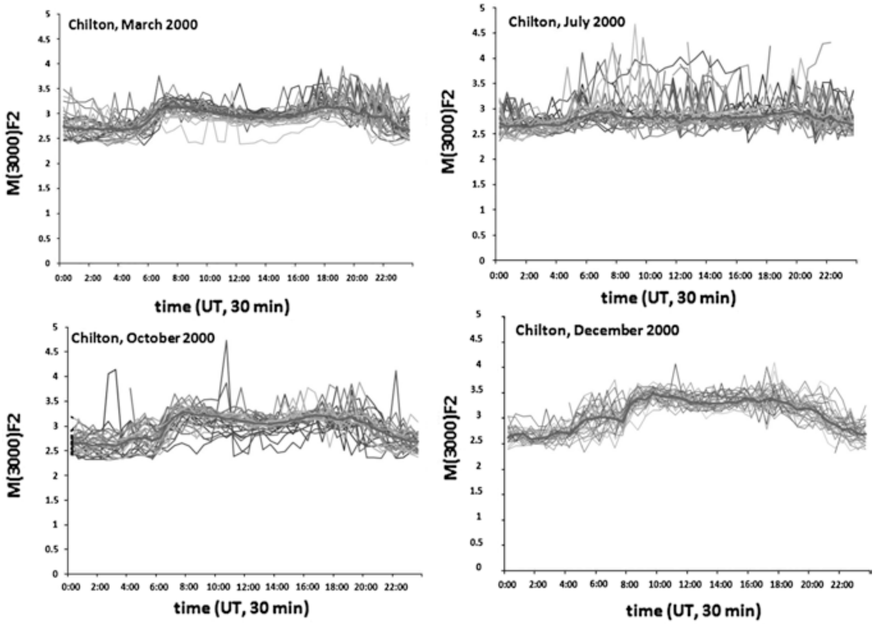


**Fig. 4.13** Variation of *M(3000)F2* at the Chilton (51.6° N, 358.7° E) ionosonde station for every day during March, July, October, and December 2000 (*bold lines* represent respective monthly median values)

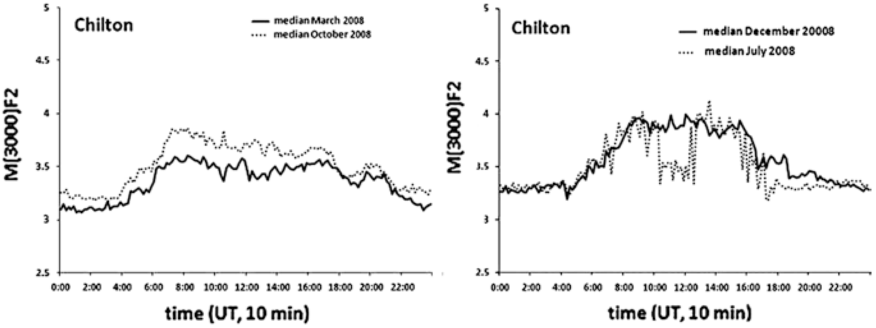


**Fig. 4.14** Diurnal variations of the *M(3000)F2* monthly median values at the Chilton (51.6° N, 358.7° E) ionosonde station during March and October 2008 (*left plot*), and July and December 2008 (*right plot*)

It is rather obvious that daytime and night-time *M(3000)F2* values have similar seasonal variability and scatter around the monthly median values. The seasonal variations of the *M(3000)F2* monthly median values during 2008 are illustrated in Fig. 4.14, which shows that the amplitude of the seasonal variation increases with decreasing solar activity.

## 4.4 Solar Cycle Variations at Mid-Latitude

One of the longest periodic variations in solar energy is a cycle averaging 11 years, corresponding to the time between reversals in the solar magnetic field. As mentioned in Sect. 2.2.3 the intensity of solar radiation for long-term prediction purposes can generally be quantified by the 12-month running average $R_{12}$ and $\Phi_{12}$ (the $\Phi$ index is often alternatively labelled *F10.7*) or the monthly mean of their daily values.

Most early studies addressed the variations of *foF2* with $R_{12}$, an index that correlated well with ionospheric changes and was proven to be a reliable means for users to establish predicted values. In the majority of these studies a near-linear variation was found up to $R_{12}$ values of around 150, when saturation sets in. Accordingly, in the ITU-R P. 1239 Recommendation on Reference Ionospheric Characteristics, the monthly medians for *foF2* and *M(3000)F2* are based on a linear relationship with solar activity. The once-and-for-all long-term *foF2* and *M(3000)F2* maps are separately formulated for $R_{12} = 0$ and 100. A linear variation is assumed up to $R_{12} = 150$ and thereafter complete saturation is assumed.

In recent years similar analysis has addressed solar cycle variations of monthly mean *foE, foF1,* and *foF2* values against monthly mean values of the solar parameters *Ri* and *F10.7*. Perhaps the best example is the solar cycle variation shown in Fig. 4.15.

Using past ionosonde measurements from Slough (51.5° N, 359.4° E), in 1994 P. A. Bradley examined the dependence of monthly median noon *foF2* and *M(3000)F2*

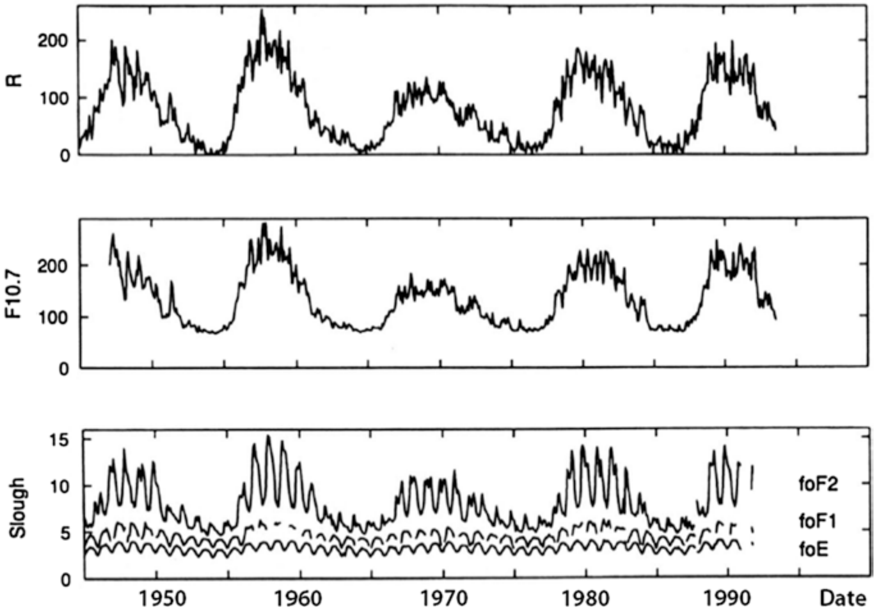


**Fig. 4.15** The noon critical frequencies *fo* (proportional to square root of peak electron densities) of the E, Fl, and F2 layers at the Slough (51.5° N, 359.4° E) ionosonde station *versus R* and *F10.7* (Courtesy Rutherford Appleton Laboratory)

on the $R_{12}$ indices over 6 solar cycles and during 4 months representing the different seasons. This statistical study was based on data for all years taken together and best-fit parabolic relationships. Figure 4.16 gives examples of the results obtained for *foF2*

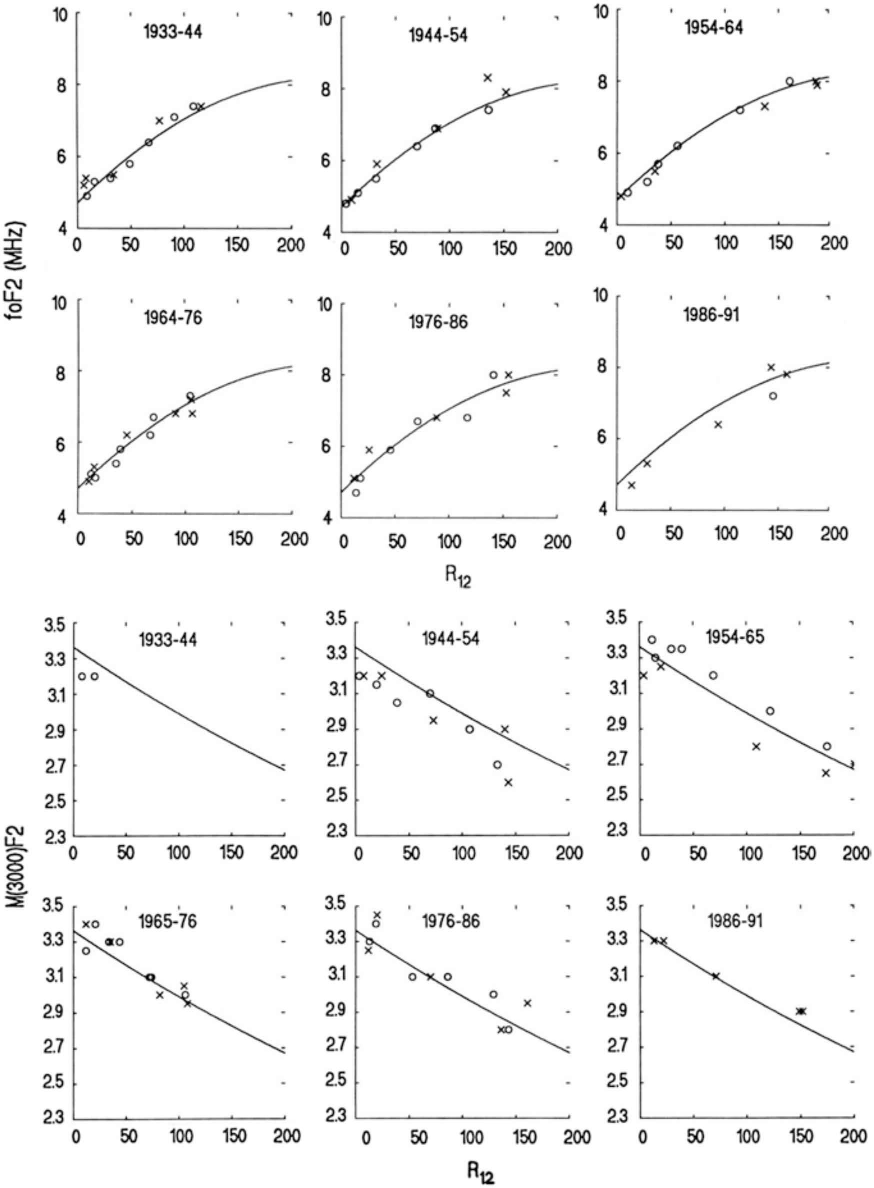


**Fig. 4.16** Examples of the solar cycle variation of noon monthly median *foF2* in June (*top panel*), and *M(3000)F2* in March (*lower panel*) with $R_{12}$ over 6 solar cycles for the Slough (51.5° N, 359.4° E) ionosonde station (from Bradley 1994)

in June, and for *M(3000)F2* in March. These results show no evidence of a long-term trend over different solar cycles. More importantly these results have confirmed parabolic solar cycle dependence of *foF2* on $R_{12}$ in the summer and equinoxes, and a linear solar cycle dependence of *foF2* on $R_{12}$ in winter and *M(3000)F2* on $R_{12}$ in all months.

Overall, the COST 238 ionospheric action PRIME, long-term mapping initiative for telecommunication purposes, concluded that a single parabolic relationship is recommended to define *foF2* solar cycle variations by combining all available measurement data regardless of which cycle they refer to, or whether during rising or falling half cycles. In the case of *M(3000)F2* a linear relationship was recommended.

However, a new related study based on ionospheric observations over Millstone Hill (42.6° N, 288.5° E), another mid-latitude ionosonde station, from an incoherent scatter radar (ISR) for two full solar cycles (1976–2002) and from an ionosonde during the periods 1989–1990 and 1998–2004, has clearly shown that solar

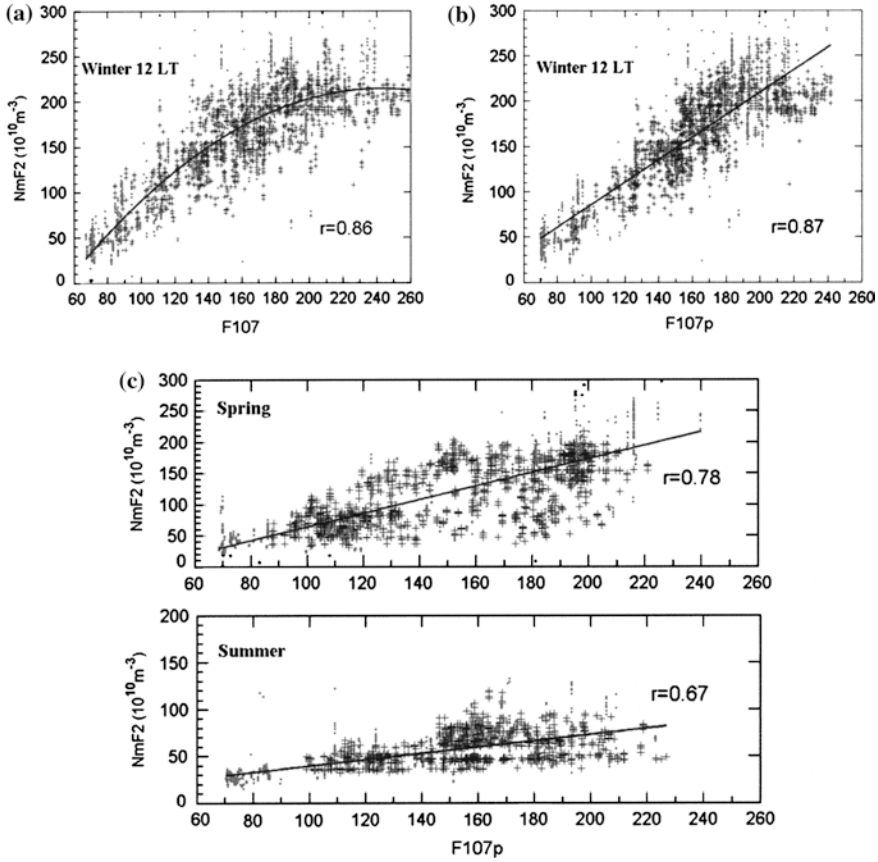


**Fig. 4.17** Relationship of the observed $N_mF2$ to the daily *F10.7* index over the Millstone Hill (42.6° N, 288.5° E) ionosonde station at 1200 LT in winter (**a**), to the solar proxy *F10.7p* index in winter (**b**), spring (**c** *top*), and summer (**c** *bottom*) (from Lei et al. 2005)

cycle low could differ between locations and for different solar indices. The results demonstrate that the observed *NmF2* (proportional to the square of *foF2*) exhibits an almost linear relationship with the solar activity index *F10.7* in all seasons. Figure 4.17a displays the response of the noon *NmF2* to the daily *F10.7* index during the winter season with the solid line representing a 2nd degree polynomial fitting and the correlation coefficient $r = 86$. By using the solar proxy index $F10.7p = (F10.7 + F10.7A)$, where *F10.7A* is the 81-day running mean of the daily *F10.7*, it is shown in Fig. 4.17b, c that a linear function can be successfully used to represent *NmF2* and *F10.7p* correlation in all seasons.

Temporal variations in the ionospheric characteristics identified here are essential for understanding the physical mechanisms involved, and for ionospheric prediction, which will be the subject of the following Chapters.

## Suggested Readings

Balan N, Shiokawa K, Otsuka Y, Watanabe S, Bailey GJ (2009) Super plasma fountain and equatorial ionization anomaly during penetration electric field. J Geophys Res. doi:10.1029/2008JA013768

Bradley PA (1994) Further study of foF2 and M(3000)F2 in different solar cycles. Ann Geofis 37:201–208

Chapman S (1931a) The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth. Proc Phys Soc 43:26–45

Chapman S (1931b) The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, II. Grazing incidence, Proc Phys Soc 43:483–501

Dow JM, Neilan RE, Rizos C (2009) The International GNSS service in a changing landscape of global navigation satellite systems. J Geod 83:191–198

Feichter E, Leitinger R (2002) Properties of the main trough of the F region derived from dynamic explorer 2 data. Ann Geofis 45:117–124

Handbook on ionospheric properties and propagation, ITU-Radiocommunication SG3, Geneva (1996)

Hargreaves JK (1979) The Upper Atmosphere and Solar-Terrestrial Relations. Van Nostrand Reinhold Co., England

ITU-R (1997) ITU-R Recommendation P. 1239, ITU-R Reference Ionospheric Characteristics. International Telecommunication Union, Geneva

Lei J, Liu L, Wan W, Zhang S-R (2005) Variations of electron density based on long-term incoherent scatter radar and ionosonde measurements over Millstone Hill. Radio Sci. doi:10.1029/2004RS003106

Mendillo M, Huang C-L, Pi X, Rishbeth H, Meier R (2005) The global ionospheric asymmetry in total electron content. J Atmos Sol-Terr Phys 67:1377–1387

Prölss GW (2006) Ionospheric F-region storms: unsolved problems. In: Characterizing the Ionosphere. Meeting Proceedings RTO-MP-IST-056, Neuilly-sur-Seine, pp 10-1–10-20

Rishbeth H (1996) Basic physics of the ionosphere. In: Hall MPP, Barclay LW, Hewitt MT (eds) Propagation of Radio waves. The Institute of Electrical Engineers, London

## Additional Web Site

http://www.rto.nato.int/abstracts.asp

# Chapter 5
# Ionospheric Models for Prediction and Forecasting

## 5.1 Introduction

A comprehensive knowledge of the Earth's ionospheric structure and dynamics, in particular its refractivity profiles, is necessary to guarantee the effective operation, planning, and management of numerous radio communication and navigation systems. These can be based either on: (1) advanced models coupled across the space-terrestrial environment, providing detailed knowledge of conditions and good predictions; or (2) specific well-targeted services that provide products like specification data, post-event analysis, long-term predictions, instantaneous nowcasting and short-term forecasting generated in real or near real-time in response to available observations.

Since its discovery, many models of the Earth's ionosphere have been developed. These fall into several classes and include the D, E, and F regions at low, middle, and high-latitudes. Ionospheric models can be divided into three main groups:

1. theoretical, parameterised, and empirical models that define not only the ionospheric electron density profile but also the different profile parameters in terms of ionospheric characteristics at every point on the globe, for example *foE, foF1*, and linear or polynomial *foF2* versus $R_{12}$;
2. assimilation models for a full 3-D electron density profile;
3. empirical and physical models or methods for 2-D global, regional, and local mapping of the characteristic ionospheric parameters both for long-term prediction and nowcasting in the field of radio propagation and navigation (e.g. CCIR and URSI maps of ITU-R, ionospheric COST actions maps, etc.).

The first two groups of models are more general and are therefore presented in this Chapter, while the third group of models and methods are more practical and are described in the following Chap. 6.

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_5,

## 5.2 The Ionospheric Inverse Problem

As described in Chap. 3, ionograms are the experimental results obtained from sweep frequency HF vertical incidence radar providing information on the state of the ionosphere directly above the ionosonde location. Apart from the ionospheric characteristic of each region, the most important scaled parameter from an ionogram is the vertical electron density profile $N\,(h)$ as a function of time. However, ionograms provide a record of the virtual height of reflection of radio waves in the ionosphere as a function of wave frequency. The relation between virtual height and real height is given by:

$$h'(f) = \int_0^{h_0(f)} \mu'(N, B, f)dh \tag{5.2.1}$$

where $h'$ and $f$ represent the virtual height of reflection and the radio wave frequency, $ho$ is the "true" or the real height, $\mu' = 1/\mu$ is the group refractive index, $N$ and $B$ are the electron density and geomagnetic field strength at a certain real height $h$.

Therefore, precision methods are required to determine the ionospheric electron density or plasma frequency profile as a function of real height from the recorded echo traces found on ionograms. These are generally known as "true height" inversion procedures of $h'\,(f)$ in Eq. 5.2.1. The inversion of data to obtain an $N\,(h)$ profile, giving electron density, is ambiguous due to the absence of direct information at low frequencies and from any valley regions between the ionospheric layers. Since the early days of ionospheric vertical incidence soundings, a variety of "true height" methods have been developed to evaluate the electron density profiles converted from ionograms. The best known is a general true height analysis program, POLAND, developed by Titheridge (1985). These profiles are then compared with those measured by incoherent scatter radar (see Sect. 3.4.1) to establish the accuracy of the inversion procedure.

The inversion procedure is generally a mathematical representation of the vertical distribution of electron density of the ionosphere in different segments. Most modern vertical incidence ionosonds are capable of autoscaling ionograms and offer the main ionospheric characteristics in real-time, including the reconstruction of an electron density $N\,(h)$ profile from the experimental data obtained from each sounding, as described in Chap. 3. However, for in-depth ionospheric research the inverse ionospheric problem and its solution remains as important an issue as the highly sophisticated ionospheric modelling illustrated here in Sects. 5.3, 5.4, and 5.5.

## 5.3 Theoretical and Parameterised Ionospheric Models

A 3-D modelling of the ionospheric electron density height profile over the whole altitude range for all geographic positions, time spans, and solar-terrestrial conditions is an essential goal of ionospheric physics and a critical element in designing

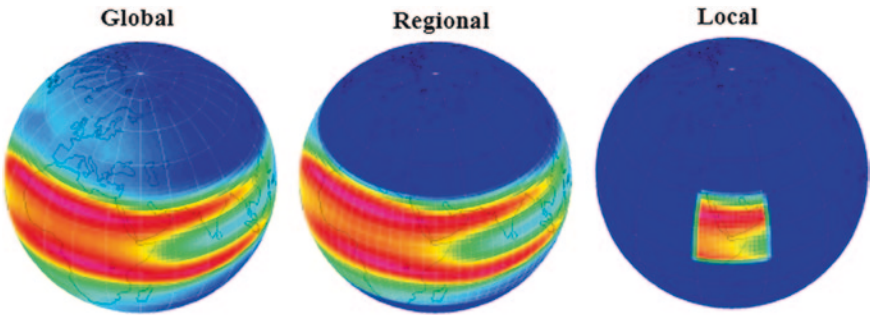


**Fig. 5.1** Ionospheric models provide both specifications and forecasts on a global, regional, or local grid

and engineering practical schemes for reliable radio communication and navigation forecasts. The existing models that can provide both specifications and forecasts on a global, regional and/or local grid are, at best, approximations of this final goal (Fig. 5.1).

Below some major aspects of existing theoretical and parametric models are described with emphasis on the key issues relevant to ionospheric prediction and forecasting. An excellent source for some of the available atmospheric, ionospheric, plasmaspheric, and other solar-terrestrial models is at http://ccmc.gsfc.nasa.gov/modelweb/.

Theoretical or physical models attempt to solve the Boltzmann equation for ionospheric plasma starting from the continuity, momentum, and energy equations for the ionized species. They calculate the ionospheric parameters from first principles using knowledge of the physical and chemical processes that control the behaviour of the magnetospheric-ionospheric-atmospheric system. As the transport equations are time-dependent, coupled, and nonlinear, different numerical methods have been used to obtain their solutions in terms of ion and electron densities, flow velocities, and temperatures as a function of altitude, latitude, longitude, and universal time. This large volume of precious information regarding ionospheric parameters is extremely important for comparisons with observed data sets and studying particular ionospheric features.

Some of the best known and widely used theoretical models are:

- Time Dependent Ionospheric Model (TDIM) developed at Utah State University, USA;
- Coupled Thermosphere-Ionosphere Model (CTIM) developed at University College London and Sheffield University, UK;
- Thermosphere-Ionosphere Global Circulation Model (TIGCM) developed at the National Center for Atmospheric Research, USA;
- Field Line Interhemispheric Plasma Model (FLIP) developed at the University of Alabama, USA; and
- Global Theoretical Ionospheric Model (GTIM) developed at Phillips Laboratory, USA.

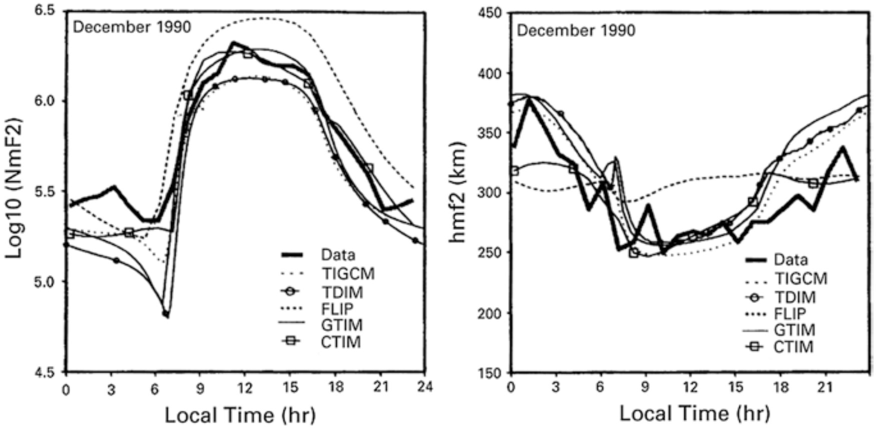


**Fig. 5.2** A comparison of diurnal variations in *NmF2* and *hmf2* observed at Millstone Hill with corresponding TDIM, CTIM, TIGCM, FLIP, and GTIM results for December 1990 (from Rodger and Jarvis 2000)

The main resource for detailed information about theoretical models is the STEP Handbook of Ionospheric Models by Schunk (1996) describing 13 of the most advanced physics-based models of the ionosphere.

Figure 5.2 shows an example of a comparison between the diurnal variations in the profile parameters *NmF2* and *hmf2* observed at Millstone Hill, derived from five theoretical ionospheric models for solar maximum winter conditions in December 1990. Even under ionospheric climatology conditions, the models tend to underestimate *NmF2* by a factor of 2 during the daytime and 10 at night.

Because of their complexity, all these models are effectively confined to modern supercomputers. Consequently, the main problem of using theoretical models for operational ionospheric prediction and forecasting is the large amount of computer time required. In addition, an extensive preparation of inputs dependent on solar, interplanetary, and geomagnetic drivers is needed to obtain meaningful results. Although, theoretical ionospheric models have proved their capabilities in reproducing certain geomagnetic storm and substorm conditions, for most operational ionospheric prediction and forecasting applications in real or near real-time this is a significant limitation.

Parameterised ionospheric models significantly simplify the theoretical models by expressing them in terms of solar-terrestrial parameters and geographical locations, giving a realistic representation of the ionospheric spatial and temporal structure using a limited number of numerical coefficients. At present there are a few parameterised versions of theoretical models available for different ionospheric regions. They are listed and briefly described in Table 5.1, which shows that such models have been developed for the low, mid, and high-latitude regions of the ionosphere.

The most popular model among those in Table 5.1 is a global climatology model called the Parameterised Ionospheric Model (PIM), obtained by combining the theoretical outputs of two previously mentioned models: the GTIM for low and mid-latitudes, and the TDIM for high-latitudes, with an empirical model

**Table 5.1** Some of the best known parameterised models of ionospheric electron density (from Cander et al. 1996)

| Model | Characteristics | Extent |
|---|---|---|
| SLIM (Anderson et al. 1987) | Based on theoretically obtained grid values for electron density profiles normalized to the F2-peak and then represented by modified Chapman function using six coefficients per individual profile | Low-latitude |
| FAIM (Anderson et al. 1989) | Uses the formalism of the Chiu model with coefficients fitted to the SLIM model profiles and requires a smoothed Zurich sunspot number | Low- and mid-latitude |
| ICED (Tascione et al. 1988) | Controlled by the sunspot number (SSN) and geomagnetic Q index and conceived to allow for real-time updates of the input parameters from a number of sensors | Global, improved performance for high-latitude |
| PIM (Daniell et al. 1995) | Amalgam of a number of other models and uses either the *foF2* CCIR coefficients for normalisation of the electron density profiles or coefficients produced by the TDIM | Global |

for plasmaspheric altitudes. The altitude electron density profiles of these models were then represented by empirical orthonormal functions (EOF) and the resulting large set of coefficients constituted the PIM model itself.

For customer specified solar-terrestrial conditions and spatial coordinates, the PIM model generates electron density profiles from altitudes of 90–25,000 km, with corresponding critical frequencies and heights for the ionospheric E and F2 layers, and ionospheric vertical total electron content (see Fig. 5.3). In many applications, PIM users also take advantage of the option to update the PIM F2 layer peak with the CCIR maps, which provide a monthly resolution instead of seasonal resolution as a standard option.

The importance of PIM also lies in the fact that a version of PIM with a capability for near real-time updates using data from ground-based digisondes and in-situ satellite measurements forms part of the first data-driven ionospheric model, called the Parameterised Real-Time Ionospheric Specification Model (PRISM). This development will lead to the novel data assimilation ionospheric modelling activities described in Sect. 5.5. The PIM and PRISM family of models were developed at the US Air Force Research Laboratories.

## 5.4 Models of the Electron Density Profile in the Ionosphere

In Sect. 2.3.3 the concept of electron density profile, or in another definition, electron concentration *versus* altitude was introduced considering its importance for both theoretical studies and practical applications.

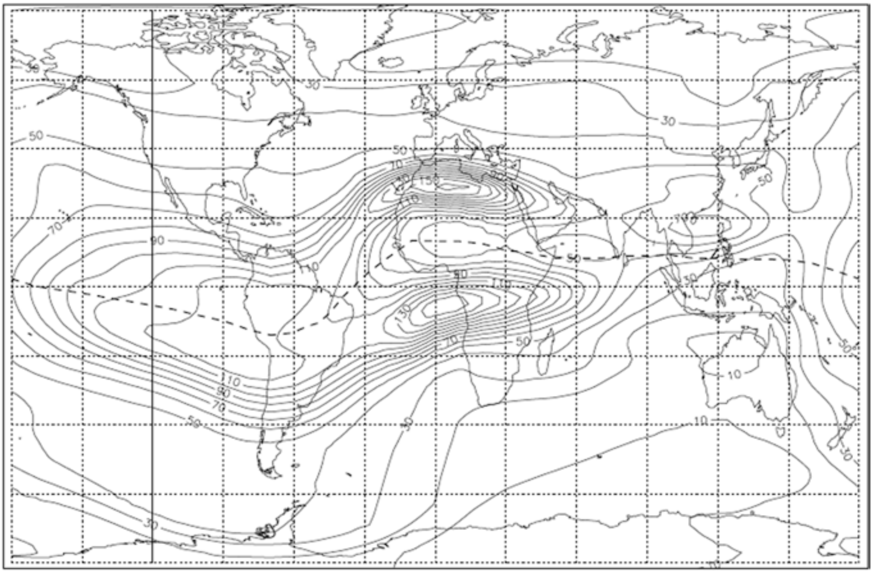


**Fig. 5.3** Vertical *TEC* calculated by PIM at 20 UT for the vernal equinox during solar maximum, in units of TECU. The longitudinal grid lines correspond to 2 h segments of local mean time; the latitudinal grid lines are at 20° intervals (from Campbell 1999)

Together with the first historical models, like the T. Shimazaki formula published in 1955 and still in use, or the simple model of P. A. Bradley and J. R. Dudeney in 1973, two analytical models now widely implemented are presented below: the International Reference Ionosphere (IRI), and the NeQuick model. These are based on analytical functions and experimental observations.

Many other models have been developed in recent years, including those in Sects. 5.3 and 5.5 like the SLIM, FAIM, and more recently the PRISM (Parameterised Real-Time Ionospheric Specification Model), capable of producing critical frequencies and peak heights for the ionospheric E and F2 layers, as well as electron density profiles from 90 to 1,600 km and *TEC*.

However, IRI and NeQuick have been widely implemented among a large number of users, geophysicists as well as technical engineers, with a easy to use FORTRAN code, or interfaced with different applications, and most importantly with continuous improvements in the methods.

### 5.4.1 True Height of Peak Density: The Shimazaki Formula and Subsequent Improvements

The first step towards developing an electron density model was to define *hm*, the real height of peak density, this being the real height of the maximum electron density of the F2 layer, using the agreed routine ionospheric characteristics obtained from ionospheric vertical sounding. The first simple empirical relation was provided by T. Shimazaki:

$$hpF2 = \frac{1{,}490}{M(3000)F2} - 176\ \text{km} \tag{5.4.1}$$

assuming that the F2 layer has an approximately parabolic shape and where *hp* is the virtual height at the frequency 0.834 *foF2*, and *M*(*3000*)*F2* is a transmission factor, one of the routinely scaled characteristics in an ionogram.

Considering that during the night *hpF2* is almost equal to *hmF2,* the height of maximum electron density, and that the difference $\Delta h$ between *hpF2* and *hmF2* is very small, the T. Shimazaki formula is now generally known in the present form:

$$hmF2 = \frac{1{,}490}{M(3000)F2} - 176\,\text{km} \tag{5.4.2}$$

even if other authors report different constants of 1,411 and 196 km instead of 1,490 and 176 km respectively.

It was only in 1973 that an important improvement was proposed by P. A. Bradley and J. R. Dudeney, introducing the following relation:

$$hmF2 = \frac{1{,}490}{M(3000)F2 + \Delta M} - 176\,\text{km} \tag{5.4.3}$$

where:

$$\Delta M = \frac{0.18}{X - 1.4} \quad \text{and} \quad X = \frac{foF2}{foE} \tag{5.4.4}$$

The parameter $X$ takes into account the modification produced by the underlying ionization present in the E region. $\Delta M$ goes to 0 and becomes similar to the old T. Shimazaki formula for high $X$ and then there no lower ionization.

Also in 1972 R. Eyfrig introduced a dependence on the index of solar activity $R_{12}$ with a new:

$$\Delta M = \frac{0.18}{X - 1.4} + 0.096\,\frac{R_{12} - 25}{150} \tag{5.4.5}$$

Finally, $\Delta M$ was further improved by D. Biltza et al. in 1979 incorporating a dependence on geomagnetic dip latitude $\varphi$, defined as:

$$\tan \Psi = \frac{1}{2} \tan \psi \tag{5.4.6}$$

with $\psi$ the inclination of the Earth's magnetic field at 300 km of altitude. So $\Delta M$, adopted since 1978 by the International Reference Ionosphere (see Sect. 5.4.3), is:

$$\Delta M = \frac{f_1 f_2}{X - f_3} + f_4 \tag{5.4.7}$$

with:

$$f_1 = 0.00232\, R_{12} + 0.222, \tag{5.4.8}$$

$$f_2 = 1 - R_{12}/150 \exp(-(\Psi/40)^2), \tag{5.4.9}$$

$$f_3 = 1.2 - 0.0116 \exp(R_{12}/41.84), \tag{5.4.10}$$

$$f_4 = 0.096(R_{12} - 25)/150. \tag{5.4.11}$$

### *5.4.2 Bradley and Dudeney Model*

A very simple model of vertical distribution of electron density was introduced by Appleton and Beynon (1947), consisting of a parabolic E region and a semi parabolic F2 layer. Following on from this, another simple model was proposed in 1973 by P. A. Bradley and J. R. Dudeney (Fig. 5.4).

This model was developed using routinely measured ionospheric characteristics like *foE, foF2*, *M(3000)F2*, or simple empirical relations for the height of the maximum electron density *hmF2* of the F2 layer, or fixed values for the height of the maximum electron density *hmE* of the E region, and *ymF2* and *ymE* for the semi thickness of the F2 layer and E region respectively. The main steps of the model are summarized as follows:

1. The E region is described with a semi parabolic distribution below the height of the maximum electron density *hmE,* here conventionally taken to be 110 km and with a semi thickness *ymE* of 20 km.

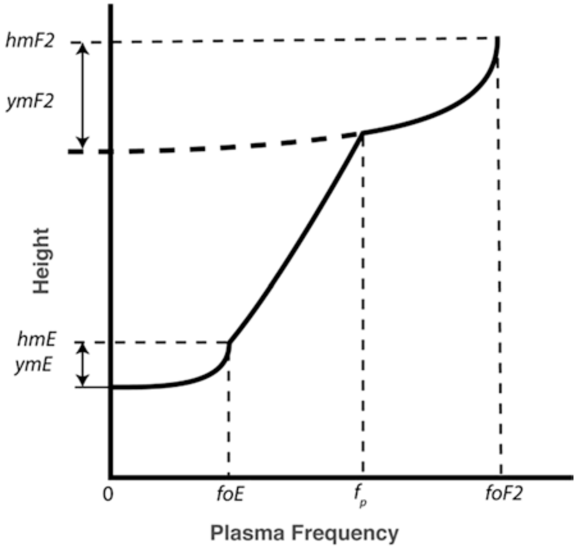


**Fig. 5.4** Scheme of the electron density profile introduced by P. A. Bradley and J. R. Dudeney; *foF2* and *foE* are routinely measured ionospheric characteristics, *hmF2* from the Shimazaki formula improved by the authors, *ymE* and *hmE* are fixed values, *ymF2* is from an empirical calculation, and *fp = 1.7 foE*

2. The F2 layer is assumed to be a semi parabolic distribution with the height of maximum electron density *hmF2* given by the empirical Eq. 5.4.3 and semi thickness *ymF2*.
3. Between the maximum height of the electron density of the E region *hmE* and the point of the parabolic distribution of the F2 layer given by the plasma frequency $fp = 1.7\ foE$, the electron density increases linearly. In this model the F1 layer characteristics are neglected, assuming that those cusps in the ionograms scaled as F1 represent only small fluctuations in electron density.
4. The semi thickness *ymF2* is given by the relation:

$$ymF2 = hmF2 - (h'F - \Delta h) \tag{5.4.12}$$

where $h'F$ is the minimum virtual height of the ordinary wave F trace routinely scaled in an ionogram, and $\Delta h$ is a correction factor given empirically as:

$$\Delta h = \left(\frac{0.618}{x - 1.33}\right)^{0.86} (hmF2 - 104) \tag{5.4.13}$$

### 5.4.3 International Reference Ionosphere (IRI): Basic Principles and Equations for the Electron Density Profile

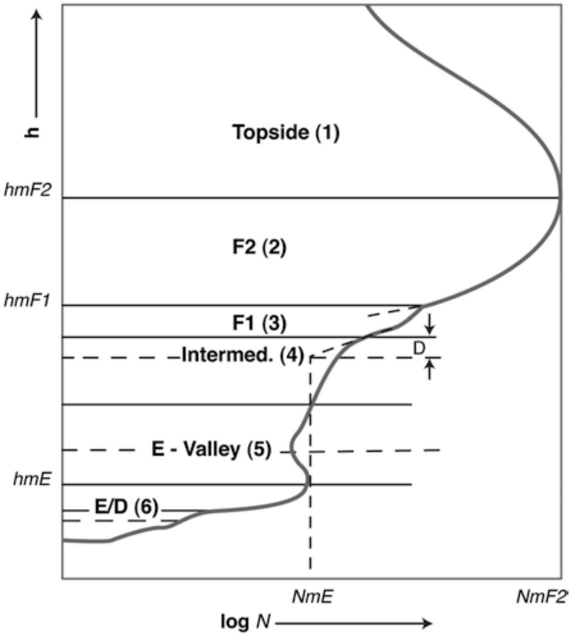
The International Reference Ionosphere is the result of the efforts of the international scientific community, who over the last 60 years have worked to improve and upgrade a standard model of the ionosphere with the support of the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI). The International Reference Ionosphere (IRI) is a complex model that describes for a given coordinate, time, and date not only the electron density in the range from roughly 50–2,000 km, but also the ion composition together with the electron and ion temperatures (http://iri.gsfc.nasa.gov/).

Over the last 50 years different and upgraded editions of IRI have been released based on the available data sources, which include the worldwide network of ionosondes, incoherent scatter radars, topside sounders, and in-situ instruments on various satellites and rockets. The great importance of the International Reference Ionosphere is in the continuous improvements and new contributions to the standard model tested and agreed by the members, together with its free distribution in FORTRAN source code.

Regarding the electron density profile, the IRI model is designed to use different models and different input data according to the required parameters obtained from analytical models or provided directly by the user. It is divided into six parts or regions as shown in Fig. 5.5, see Bilitza (1990).

1. The topside of the F region.
2. The bottom-side of the F2 layer.

**Fig. 5.5** Scheme of the IRI electron density profile divided into six regions (from Bilitza 1990)



3. The F1 layer.
4. The intermediate region.
5. The valley of the E region.
6. The bottom-side of the E region and the D region.

The borders of these divisions are recognized by characteristic points on the electron density profile coincident with the relative maxima of density over the different F2, F1, and E layers. These maxima, also defined as critical frequencies *foF2*, *foF1*, and *foE*, are usually still monitored by the worldwide network of vertical ionospheric stations. These characteristics are related to electron density in the equation seen in Chap. 3:

$$\mathrm{Nm/m^{-3}} = 1.24\ 10^{10}\ (\mathrm{f/MHz})^2 \tag{5.4.14}$$

#### 5.4.3.1 The Topside of the F Region

The topside of the F region is the section of the ionosphere from the height *hmF2* of maximum electron density extending up to 1,000 km in altitude, or up to the beginning of the plasmasphere. This part of the ionosphere is particularly important for evaluating the *TEC*, the total electron content, because of its role in satellite positioning and navigation systems. The behaviour of electron density in this region was first approximated by a Chapman layer and described by a monotonic decreasing function. To avoid the over estimation of electron density generated by the original IRI model, due to the limited database of topside soundings or in-situ measurement by incoherent scatter radar, two options were recently introduced to improve the performance of the model to altitudes above the F2 peak. The first consists in applying a correction factor based on a large number of measurements from the satellites Alouette 1 and ISIS 1 and 2 as a function of altitude, modified latitude, and local time. The second option was introduced by the NeQuick model, described in Sect. 5.4.4, using a semi Epstein's function with a height dependent thickness parameter calculated by fitting a large number of topside profiles from the Alouette and Intercosmos satellites.

Different options are provided in the IRI model to evaluate the characteristic parameter *NmF2* in relation to the critical frequency *foF2*.

Firstly, the user may include as input data, measured values of ionospheric characteristics from an ionosonde located close to the coordinate considered.

Secondly, climatological models can be used to obtain median monthly prediction values of *foF2* and *M(3000)F2* over all points of the Earth for every season, hour of day, and solar cycle activity phase.

Within the IRI software code the global models and respective numerical coefficients are based on the mapping methods of the CCIR and URSI global models described in detail in Chap. 6. These are based on a set of observations of monthly median values from the worldwide network of ionosondes over a period of years and embracing different indices of solar activity.

The performance of the model depends on the capacity of the mapping model to predict or to make a retrospective evaluation of parameters for a point of generic coordinates, above land or even the ocean. However, it is important to note that for global models, unlike mid-latitude regional or local models, the influence of the magnetic field on the F region is significant, and so geomagnetic coordinates are used rather than geographic coordinates. Then good results have been obtained by introducing the modified magnetic dip latitude $\mu$ or "modip" given by the following relation:

$$\tan \mu = \psi / \cos^{1/2} \varphi \qquad (5.4.15)$$

with $\psi$ the magnetic dip defined by Eq. 5.4.6 and $\varphi$ the geodetic latitude.

The T. Shimazaki formula, improved by P. A. Bradley and J. R. Dudeney, R. Eyfrig, and D. Biltza, was adopted by IRI to evaluate the height of maximum electron density *hmF2* (see Eqs. 5.4.3, 5.4.5, and 5.4.7).

The shape of the profile based on satellite measurements is analytically represented by an Epstein function as:

$$N(h)/NmF2 = \exp(-1/\alpha \cdot B(x - x_0, 2, TG, TX, TC)) \qquad (5.4.16)$$

Here $B$ it is the Booker function and $x - x_0$ is the modified altitude according to the relation:

$$x - x_0 = \alpha(h - hmF2) \quad \alpha = 700/(1000 - hmF2) \quad \text{and} \quad x_0 = 300 - \delta \qquad (5.4.17)$$

where *TG, TX, TC* are tabulated constants and $\delta$ is a corrective term.

#### 5.4.3.2 The Bottom-Side of the F2 Layer

A description of the bottom-side shape of the F2 region was defined by Ramakrishnan and Rawer (1972) with the relation:

$$N\,(h)\,/NmF2 = \exp\left(-x^{B1}\right)/\cosh(x) \text{ with } x = (hmF2 - h)/B_0 \quad (5.4.18)$$

Here $B1$ in most cases is equal to 3 while for $B_0$ there are many options of tabulated values based on empirical models. There is still no analytical function for these two parameters and during recent years different tabulated values have been proposed of variable performance when applied to different cases (see Fig. 5.6).

#### 5.4.3.3 The F1 Layer

The F1 Layer can clearly be considered as a Chapman layer and so the model introduced by Ducharme et al. (1971) is still satisfactory. The relation of the critical function *foF1* of this layer and the index of solar activity $R_{12}$, the geomagnetic dip latitude $\Psi$, and the solar zenith angle $\chi$ is:

$$foF1 = fs\,\cos^n\chi \quad (5.4.19)$$

with:

$$fs = fo(f_{100} - fo)R_{12}/100$$

$$fo = 4.35 + 0.058\,|\Psi| - 0.00012\Psi^2$$

$$f_{100} = 5.35 + 0.011\,|\Psi| - 0.00023\,\Psi^2$$

$$n = 0.093 + 0.0046\,|\Psi| - 0.000054\,\Psi^2 + 0.0003R_{12}$$

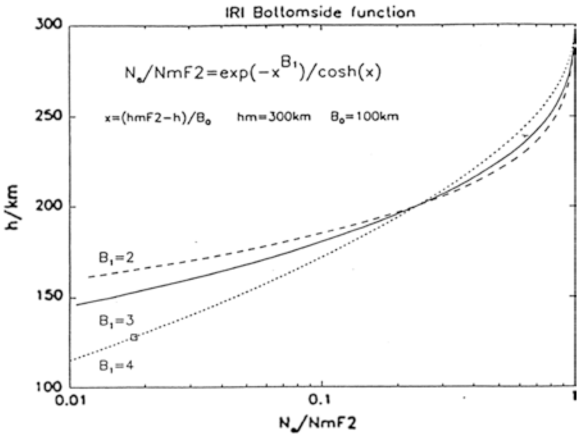


**Fig. 5.6** Different bottom-side behaviours for different values of *B1* (from Bilitza 1990)

Obviously the IRI model does not consider any occurrence of the F1 layer during the night or winter months. A parameter providing the probability of occurrence of the F1 layer was added to the model by Scotto et al. (1998). However, if the F1 layer exists, the shape of the profile can again be described by an equation as:

$$N(h)/NmF2 = \exp\left(-x^{B1}\right)/\cosh(x) + C_1\left((hmF1 - h)/B_0\right)^{1/2} \quad (5.4.20)$$

with $C_1$ a constant calculated for two different values of modified dip latitude <18° or other value.

#### 5.4.3.4 The E Region

Like the F1 layer, the E region can also be considered a Chapman layer. The critical frequency of E for this layer was provided by the model of Kouris and Muggleton (1973), depending on a radio flux index of 10.7, with the solar zenith angle $\chi$ and geographic latitude $\varphi$:

$$foE^4 = A\,B\,C\,D \quad (5.4.21)$$

with:

$$A = 1 + 0.0094\,(F10.7_{12} - 66)\,;\ B = \cos^m \chi_{noon}$$

$$m = -1.93 + 1.92\cos\varphi \ \text{ for } \ |\varphi| < 32^0 \text{ and } m = 0.11 - 0.49\cos\varphi \ \text{ for } \ |\varphi| \geq 32^0$$

$$C = 23 + 116\cos\varphi \quad \text{for} \quad |\varphi| < 32^0 \quad \text{and} \quad C = 92 + 35\cos\varphi \quad \text{for} \quad |\varphi| \geq 32^0$$

$$D = \cos^n \chi \quad \text{and} \quad n = 1.2 \quad \text{for} \quad |\varphi| > 12^0 \quad \text{and} \quad n = 1.31 \quad \text{for} \quad |\varphi| \leq 12^0$$

The behaviour of the valley between the F1 and the E region is described by a series as:

$$N(h)/NmE = 1 + E_1x^2 + E_2x^3 + E_3x^4 + E_4x^5 \quad \text{with} \quad x = h - hmE \quad (5.4.22)$$

where the values $E_1$, $E_2$, $E_3$, and $E_4$ are calculated empirically on the base of incoherent scatter measurements.

#### 5.4.3.5 The D Region

Finally, for the D region an empirical model based on in-situ measurements obtained from rocket launches was adopted:

$$NmD/\left(10^8 m^{-3}\right) = (6.05 + 0.088R_{12})\exp(-0.1/\cos^{2.7}\chi) \quad (5.4.23)$$

a function of the solar index $R_{12}$ and the solar zenith angle $\chi$; the shape of the profile is given by a third order polynomial similar to that for the E region:

$$N\left(h\right)/NmD = \exp\left(F_1x + F_2x^2 + F_3x^3\right) \quad \text{with} \quad x = h - hmD \tag{5.4.24}$$

where the $F$ parameters are obtained by in-situ measurements. To connect the D region to the E region the following exponential function is applied:

$$N\left(h\right)/NmE = \exp(-D_1\left(hmE - h\right)^K) \tag{5.4.25}$$

with $D_1$ and $K$ empirically determined to obtain a smoothed connection between the two regions.

### 5.4.4 The NeQuick Model

NeQuick, described in detail by Hochegger et al. (2000), Radicella and Leitinger (2001), is a three dimensional and time dependent ionospheric electron density model developed at the Aeronomy and Radiopropagation Laboratory of the Abdus Salam International Centre for Theoretical Physics (ICTP)—Trieste, Italy and at the Institute for Geophysics, Astrophysics and Meteorology of the University of Graz, Austria.

It is an evolution of the DGR model developed by Di Giovanni and Radicella (1990) and later improved by Radicella and Zhang in 1995. It gives an analytical representation of the vertical profile of electron density, continuous with first and second derivatives. It is a quick-run model particularly tailored for trans-ionospheric applications that allows calculation of the electron concentration at any given location in the ionosphere and thus the Total Electron Content (*TEC*) along any ground-to-satellite ray path by means of numerical integration.

The first version of the model was used by the European Space Agency (ESA) European Geostationary Navigation Overlay Service (EGNOS) project for assessment analysis and has been adopted for single-frequency positioning applications in the framework of the European Galileo project. It has also been adopted by the ITU-R as a suitable method for total electron content modelling. Several efforts, directed toward the development of a new version of the model, have recently led to the implementation of the NeQuick 2 model version. Its full analytical formulation is given in Nava et al. (2008).

The NeQuick model in all its versions uses simple analytical expressions essentially based on the Epstein layer introduced by Rawer (1963)

$$N(h; h\max, N\max, B) = \frac{4N\max}{\left(1 + \exp\left(\frac{h - h\max}{B}\right)\right)^2} \exp\left(\frac{h - h\max}{B}\right) \tag{5.4.26}$$

where *Nmax* is the maximum electron density of the layer, *hmax* its altitude, and *B* a layer thickness parameter.

The bottom-side in the most recent version of the NeQuick model is expressed by the sum of semi-Epstein layers:

$$N_{bot}\left(h\right) = N_{E}\left(h\right) + N_{F1}\left(h\right) + N_{F2}\left(h\right) \tag{5.4.27}$$

The thickness parameters take different values for the bottom-side and for the topside of each layer (BEbot and BEtop for the E layer, B1bot and B1top for the F1 layer, B2bot for the F2 layer).

The model topside is represented by a semi-Epstein layer with a height-dependent thickness parameter $H$:

$$N(h) = \frac{4NmF2}{\left(1 + \exp\left(z\right)\right)^{2}} \exp\left(z\right) \tag{5.4.28}$$

with:

$$z = h - \frac{hmF2}{H} \tag{5.4.29}$$

where:

$$H = kB2_{bot}\left[1 + \frac{rg(h - hmF2)}{rkB2_{bot} + g(h - hmF2)}\right]$$

with parameter $k$ defined as:

$$k = 3.22 - 0.0538foF2 - 0.00664hmF2 + 0.113\frac{hmF2}{B2_{bot}} + 0.00257R_{12}$$

and r = 100, g = 0.125 as constants.

Figure 5.7 shows NeQuick electron density profiles along paths at different elevation angles.

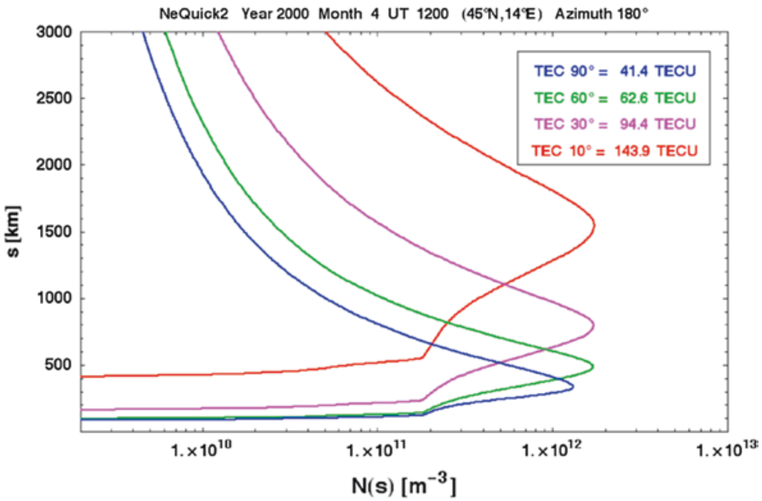


**Fig. 5.7** Example of NeQuick2 profiles and TEC along ray paths. Different colours correspond to different path elevation angles. 90° means vertical profile and *TEC* (from Radicella 2009)

## 5.5 Assimilation Models in the Ionosphere

Some advanced models are now being developed with a major impact on the expansion of ionospheric operational services. These are physics-based data-driven models that apply data assimilation techniques to specify ionospheric plasma distributions. The terrestrial weather community and oceanographers have been using such techniques for decades, with the assimilation of large numbers of observations and the solution of coupled physical equations on very large computer systems.

Currently, the best known ionospheric model of this kind is the Utah State University Global Assimilation of Ionospheric Measurements (GAIM), which is a physics-based data-driven assimilation model of the ionosphere and neutral atmosphere.

GAIM is a four-level system providing time-dependent electron density *Ne* distributions using: (1) Time-dependent climatology from the physics-based Ionospheric Forecast Model (IFM) which relies on several empirical models of the thermosphere, thermospheric winds, and electric fields; (2) An observational database to adjust the empirical drivers so that they are consistent with measurements, then running the Ionosphere–Plasmasphere Model with the adjusted drivers for simulation; (3) Kalman filters that combine the simulation results with the available real-time data; and (4) GAIM forecast mode which provides specifications and forecasts for 3-D *Ne* distribution from 90 to 25,000 km on a global, regional, and local grid depending on the operational demands and available data sets.

It is important to emphasize that GAIM assimilates from multiple data sources exactly as they are measured in real-time or near real-time: (1) bottomside *Ne* profiles from Digisondes; (2) slant *TEC* from ground-based GPS receivers; (3) *Ne* along satellite tracks; (4) integrated UV emissions; and (5) occultation data. Figure 5.8 shows contour plots of plasma densities matching the weather (TRUTH) simulation, the climate run GAIM 0 model, the “best-guess” ionosphere GAIM 1 model, and the Kalman filter reconstruction GAIM 2 model together with height profiles of the plasma density at the magnetic equator for these four cases. It is clear that in this case Kalman filter analysis by the GAIM 2 model along a geomagnetic meridional plane from 100 km up to 1,500 km in altitude after 10 h of assimilation time was successful in capturing the ionospheric density distribution. Since the results correspond to the Peruvian sector at 2200 LT on the first day of assimilation, that is 335 in 1998, it is important to note that this filter was able to track both the equatorial layer height and the strength of the equatorial anomaly.

In addition, the latest version of GAIM provides global distributions for the ionospheric drivers (neutral winds, electric fields, and particle precipitation), and quantitative estimates for the accuracy of the reconstructed ionospheric densities.

The Electron Density Assimilative Model (EDAM), from QinetiQ in the UK, was developed to provide real-time characterizations of the ionosphere by assimilating diverse data sets into a background model in order to produce a 3-D representation of electron density.

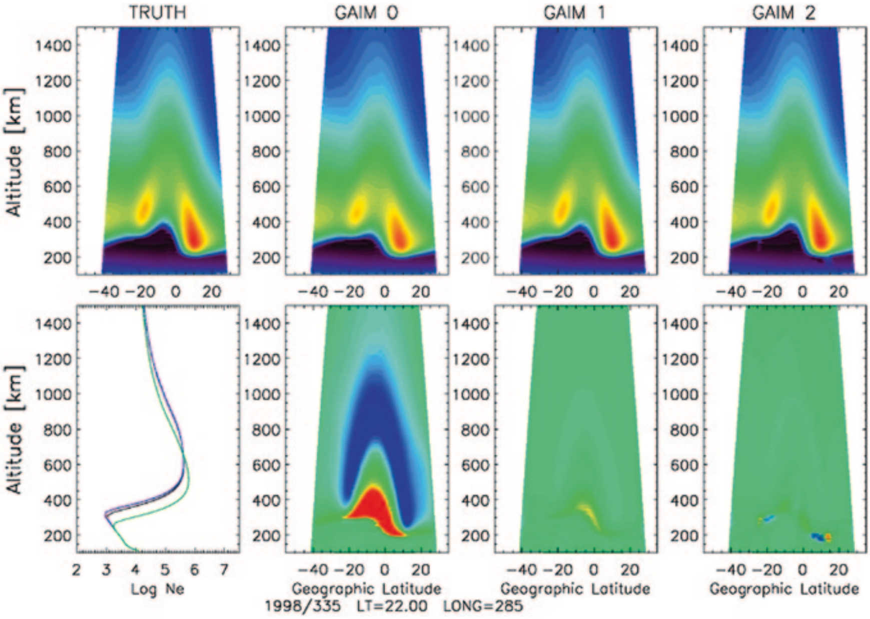


**Fig. 5.8** Contour plots of plasma densities in the *upper panel* (colour-coded from $10^3$ cm$^{-3}$ in *blue* to $10^7$ cm$^{-3}$ in *purple*) correspond to the TRUTH simulation, the GAIM 0, GAIM 1, and GAIM 2 models. Height profiles of the plasma density at the magnetic equator for these four cases and percentage differences between respective models and TRUTH simulation are shown in the *lower panel* from *left* to *right*. A *red* and *blue* colour in the differences represents an overestimation or underestimation of the "true" densities, respectively (from Scherliess et al. 2004)

EDAM is a compromise solution for the ionospheric imaging problem lying between full physical model assimilation systems and more data driven tomographic solutions. It makes use of the median PIM model for its background, with Gauss-Markov Kalman filter assimilation. Although EDAM is capable of ingesting integrated *TEC* data from radio occultation instruments as well as in-situ measurements of electron density, it is usually run with input data from the GNSS network stations and from the worldwide vertical ionosonde networks (Fig. 5.9).

It has been found that such an assimilative model can be successful in real-time estimations of HF propagation conditions. However, it is understandable that its performance is limited by the accuracy of the bottom-side electron density profile representation. Therefore, once again the ionosonde networks as data sources play a key role in constraining the bottom-side shape.

Furthermore, on the assumption of space-sparse ionospheric measurements, data assimilation is a process of merging measurement data with a model to estimate the ionospheric conditions over an area where direct measurements are not available. By means of data assimilation, it is possible to expand the effectiveness of limited measurements using the model and, at the same time, to increase the accuracy of model estimates using the measurements.

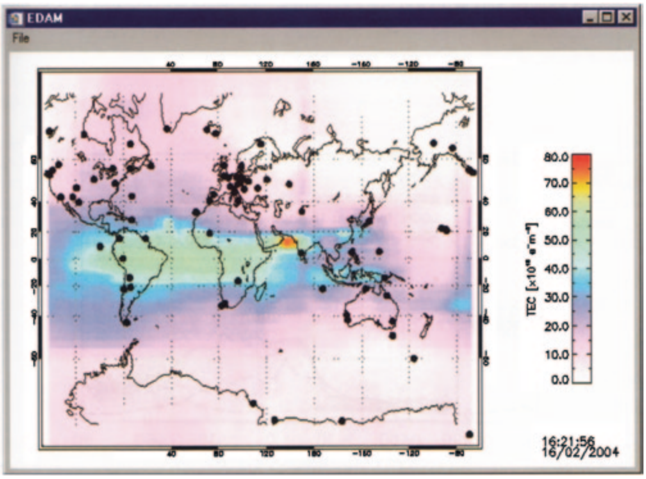


**Fig. 5.9** Screen grab showing vertical *TEC* generated from the EDAM electron density analysis. The *dots* show the locations of contributing IGS stations (from Jakowski et al. 2004)

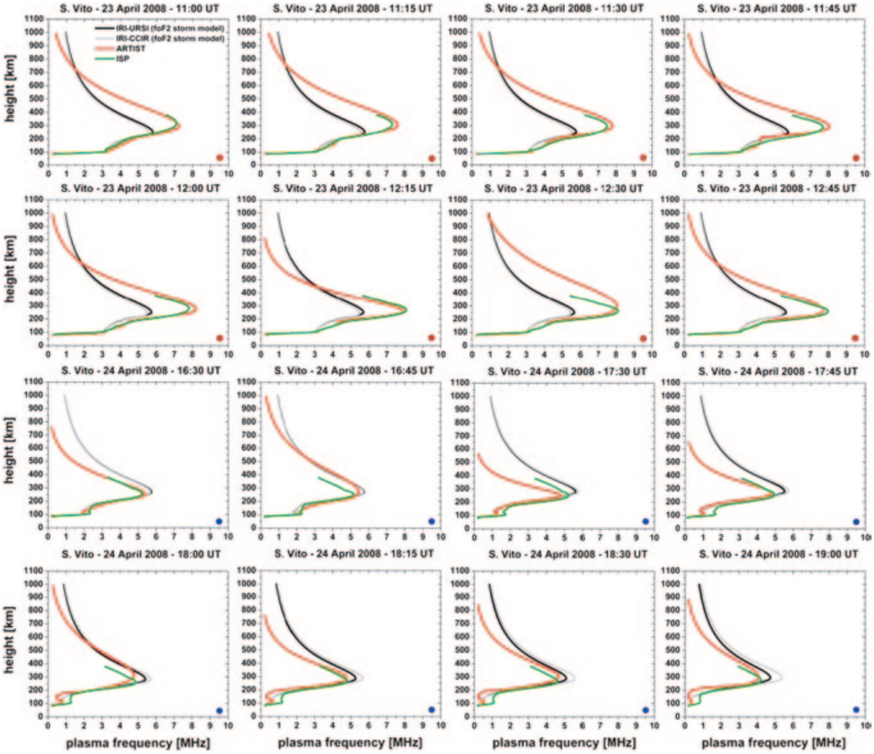


**Fig. 5.10** Comparison between the profiles obtained at San Vito (40.6° N, 17.8° E) on 23 April 2008 from 1100 to 1900 UT by IRI-SIRMUP-P (ISP, *green*), IRI-URSI (*black*), and IRI-CCIR (*gray*) (from Pezzopane et al. 2013)

For this reason, it is important to note a new model developed at INGV in Italy by Pezzopane et al. (2011) for which updated values of *foF2* and *M(3000)F2*, coming from the regional ionospheric nowcasting model SIRMUP, plus the entire electron density profiles from the autoscaled inversion of the ionograms recorded at selected reference stations, are used as input sources for real-time 3-D IRI modelling. This new approach is applied to estimate the electron density on a regional grid of the ionosphere in the Central Mediterranean area, extending in latitude from 30 to 44° N and in longitude from 5° W to 40° E with a 1° × 1° resolution. Figure 5.10 shows an example of this 3-D model validation by comparing the corresponding electron density profiles with those directly measured at one of the ionospheric test stations, in this case at San Vito (40.6° N, 17.8° E).

Finally, in order to be effective, ionospheric prediction and forecast models must add value to operational missions. Therefore, the most obvious challenges to assimilation are related to the large and disparate data sets to assimilate, the model drivers (like electric fields and wind), the integrated quantities (e.g. *TEC* and radiance), the multiple sensors, and the sub-model that needs to be included. The accuracy of the physics-based and empirical models together with the correct forecasting of drivers is critical for any operational procedure. Although assimilation models are already available, like the three briefly described above, further development is expected in coming years.

## Suggested Readings

Anderson DN (1993) Global ionospheric modelling. In: Matsumoto H (ed) Modern Radio Science, Oxford University Press, Oxford

Anderson DN, Mendillo M, Herniter B (1987) A semi-empirical low-latitude ionospheric model. Radio Sci 22:292–306

Anderson DN, Forbes JM, Codrescu M (1989) A fully analytical, low- and middle-latitude ionospheric model. J Geophys Res 94:1520–1524

Angling MJ, Jackson-Booth NK (2011) A short note on the assimilation of collocated and concurrent GPS and ionosonde data into the electron density assimilative model. Radio Sci. doi: 10.1029/2010RS004566

Angling MJ, Khattatov B (2006) Comparative study of two assimilative models of the ionosphere. Radio Sci. doi:10.1029/2005RS003372

Appleton EV, Beynon WJG (1947) The application of ionospheric data to radio communications problems, part II. Proc Phys Soc 59:58–76

Bilitza D (1990) International Reference Ionosphere 1990. Report NSSDC/WDC-R&S 90-22, National Space Science Data Center, WDC A for Rockets and Satellites, Lanham

Bilitza D (2001) International Reference Ionosphere 2000. Radio Sci 36:261-275

Bilitza D (2002) Ionospheric models for radio propagation studies. In: Stone WR (ed) The Review of Radio Science 1999–2002: Advances in 3G Mobile Communications, Cryptography and Computer Security, EMC for Integrated Circuits, Remote Sensing, Radio Astronomy and More, IEEE Press, Piscataway, NY

Bilitza D, Reinisch BW (2008) International reference ionosphere: improvements and new parameters. Adv Space Res 42:599–609

Booker HC (1977) Fitting of multiregion ionospheric profiles of electron density by a single analytic function of height. J Atmos Terr Phys 39:619–623

Bradley PA, Dudeney JR (1973) Vertical distribution of electron concentration in the ionosphere. J Atmos Terr Phys 35:2131–2146

Campbell RM (1999) Ionospheric corrections via PIM and real-time data. New Astron Rev 43:617–621

Cander LjR (2008) Ionospheric research and space weather services. J Atmos Sol-Terr Phys 70:1870–1878

Cander LjR, Zolesi B, Bradley PA (1996) Status of available N (h) model profiles. Ann Geofis 39:727–733

Daniell RE, Brown LD, Anderson DN, Fox MW, Doherty PH, Decker DT, Sojka JJ, Schunk RW (1995) Parameterized ionospheric model: A global ionospheric parameterization based on first principle models. Radio Sci 30:1499–1510

Di Giovanni G, Radicella SM (1990) An analytical model of the electron density profile in the ionosphere. Adv Space Res 10:27–30

Ducharme ED, Petrie LE, Eyfrig R (1971) A method for predicting the F1 layer critical frequency. Radio Sci 6:369–378

Gulayeva TL (1987) Progress in ionospheric informatic based on electron density profile analysis of ionograms. Adv Space Res 7:39–48

Hochegger G, Nava B, Radicella SM, Leitinger R (2000) A family of ionospheric models for different uses. Phys Chem Earth 25:307–310

Jakowski N, Leitinger R, Angling M (2004) Radio occultation techniques for probing the ionosphere. Ann Geofis 47:1049–1066

Kouris SS, Muggleton LM (1973) Diurnal variation in the E layer ionization. J Atmos Terr Phys 35:133–139

Nava B, Coisson P, Radicella SM (2008) A new version of the NeQuick ionosphere electron density model. J Atmos Sol-Terr Phys. doi:10.1016/j.jastp.2008.01.015

Pezzopane M, Pietrella M, Pignatelli A, Zolesi B, Cander LjR (2011) Assimilation of autoscaled data and regional and local ionospheric models as input sources for real-time 3-D international reference ionosphere modeling. Radio Sci. doi:10.1029/2011RS004697

Pezzopane M, Pietrella M, Pignatelli A, Zolesi B, Cander LjR (2013) Testing the three-dimensional IRI-SIRMUP-P mapping of the ionosphere for disturbed periods. Adv Space Res (in press)

Radicella SM (2009) The NeQuick model genesis uses and evolution. Ann Geophys 52:417–422

Radicella SM, Leitinger R (2001) The evolution of the DGR approach to model electron density profiles. Adv Space Res 27:35–40

Radicella SM, Zhang ML (1995) The improved DGR analytical model of electron density height profile and total electron content in the ionosphere. Ann Geofis 38:35–41

Ramakrishnan S, Rawer K (1972) Model electron density profiles obtained by empirical procedure. Space Res. XII, Academic Verlag, Berlin

Rawer K (1963) Propagation of decameter waves (HF band). In: Landmark B (ed) Meteorological and Astronomical Influences on Radio Wave Propagation, Academic, New York

Rawer K, Eyfrig R (2004) Improving the M (3000)-hmF2 relation. Adv Space Res 33:878–879

Rodger AS, Jarvis MJ (2000) Ionospheric research 50 years ago, today and tomorrow. J Atmos Sol-Terr Phys 62:1629–1645

Scherliess L, Schunk RW, Sojka JJ, Thompson DC (2004) Development of a physics-based reduced state Kalman filter for the ionosphere. Radio Sci 39 RS1S04. doi:10.1029/2002RS002797

Schunk RW (ed) (1996) Handbook of Ionospheric Models, Special Report, Boulder, USA, Solar-Terrestrial Energy Program (STEP), Scientific Committee on Solar-Terrestrial Physics (SCOSTEP)

Schunk RW, Scherliess L, Sojka JJ, Thompson DC, Anderson DN, Codrescu M, Minter C, Fuller-Rowell TJ, Heelis RA, Hairston M, Howe BM (2004) Global Assimilation of Ionospheric Measurements (GAIM). Radio Sci. doi:10.1029/2002RS002794

Scotto C, Radicella SM, Zolesi B (1998) An improved probability function to predict the F1 layer occurrence and L condition. Radio Sci 33:1763–1765

Shimazaki T (1955) Word-wide daily variations in the height of the maximum electron density of the ionospheric F2 layer. J Radio Res Lab Jpn 2:85–97

Sojka JJ (1989) Global scale, physical models of the F region ionosphere. Rev Geophys 27:371–403

Tascione TF, Kroehl HW, Creiger R, Freeman JW, Wolf RA, Spiro RW, Hilmer RV, Shade JW, Hausman BA (1988) New ionospheric and magnetospheric specification models. Radio Sci 23:211–222

Titheridge JE (1985) Ionogram analysis with the generalized program POLAN. Rep UAG-93, World Data Center A for Solar-Terr. Phys., NOAA, Environmental Data Service, Asheville

# Chapter 6
# Ionospheric Prediction for Radio Propagation Purposes

## 6.1 Introduction

Prediction of ionospheric conditions through time has always been a component of ionospheric services. Fundamental for F region modelling is the well documented fact that the critical frequency of the F2 layer *foF2* and its propagation factor *M(3000)F2*, depend systematically on measurable quantities closely related to solar radiation (Chaps. 2 and 4). Solar and ionospheric indices generated from both solar and ionospheric data provide a quick and proxy measure of the complex ionospheric response to variations in solar activity, and for this reason they are suitable for mapping the temporal variability of the ionospheric characteristics relevant to propagation prediction. The F1 and E layers have a less complex morphology and Chapman's approximations (Chaps. 2 and 4) provide an adequate temporal and spatial description for the same purposes.

Systematic description of ionospheric conditions in space using a geographical mapping method was initiated in response to the need to establish long distance HF communications, which requires knowledge of ionospheric layer reflectivity. Increasing military demands for radio communications during the last century, and in particular during the Second World War, resulted in ionospheric vertical sounders becoming sufficiently widespread to establish numerical methods for mapping the principal ionospheric characteristics including *foF2* and *M(3000)F2*. The historical evolution of ionospheric mapping and prediction methods before and just after the Second World War was described by Rawer (1975), taking into account the different points of view of the conflicting parties. There was subsequently further study and development of global mapping methods both in time and space in response to wider applications that went far beyond long distance HF radio links.

In relatively recent times there has been a trend to establish regional methods that offer better results when interfaced with other geophysical models, both for telecommunications and geophysical modelling. The proposal for an organic study of different regional techniques in the European area was an important element of the COST 238, 251, and 271 actions. During these actions numerous regional

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_6,

prediction and mapping methods were developed, examined, and tested for application in Europe and other areas, and finally compared with the performance of the most widely used global method: the CCIR/URSI-method here described in Sect. 6.2.

The study and application of regional techniques for median conditions as well as for both instantaneous and real-time specification (Sects. 6.3 and 6.4), arose from the need to improve performance and in response to the availability of a more dense network of stations. The aim was also to simplify the complex ionospheric morphology over a restricted area. However, it is not completely clear whether a regional technique will always give better results than a global one. Regional methods are very sensitive to incorrect use, and wrong or inadequately validated data banks, and can lead to an over exaggeration of local variations beyond real behaviour, as for example in the polar regions.

Finally, it is important to note that satellite observations and in-situ measurements of ionospheric parameters have not been discussed as regards improvements in global mapping. Solar activity related effects, like saturation and hysteresis, are beyond the immediate scope of the techniques presented here as relevant to ionospheric prediction for radio propagation purposes.

## 6.2 Long-Term Prediction Maps

### *6.2.1 Global Prediction Maps*

The systematic collection of routine observations from numerous ground-based stations just before the International Geophysical Year (IGY) in 1958, as well as the evolution of the first computing systems, allowed W. B. Jones and R. M. Gallet in 1960 and 1965 to develop the first important method to globally map the two key characteristics for radio propagation: *foF2* and *M(3000)F2*. Since 1983 the method has been significantly improved to avoid several problems resulting from a lack of observations over the oceans, and the inconsistent quality of the data from different ionospheric observatories. The mathematical aspect has remained the same, representing a reference method for comparison with new methods. Of course, the continuous evolution of *calculus* systems from large to personal computers has made it much easier to apply the method.

#### 6.2.1.1 CCIR *foF2* and *M(3000)F2* Prediction Maps

The CCIR (International Radio Consultative Committee) was first to provide expressions, numerical coefficients, and software that can be run on personal computers to map, in universal time (UT), the principal ionospheric characteristics used to predict radio propagation frequencies, valid for any location on the Earth and for any month,

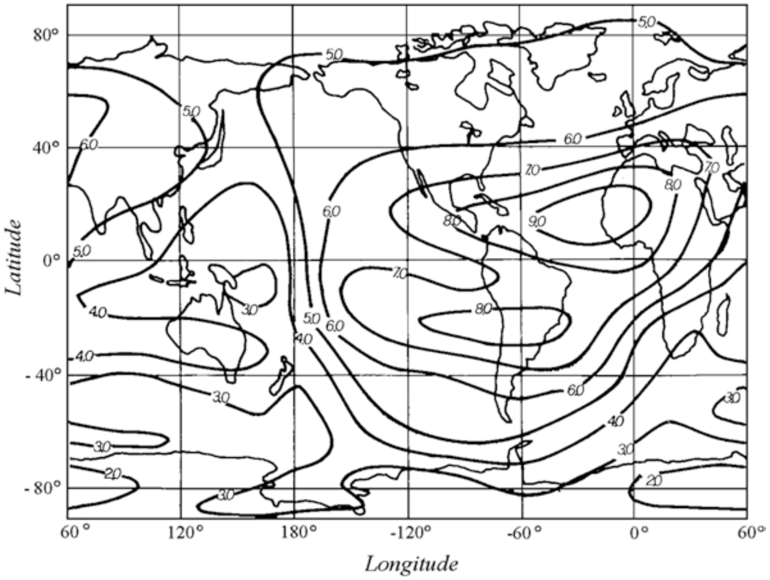


**Fig. 6.1** An example of predicted median *MUF(0)F2* map for June 1963 at 1800 UT (values are given in MHz)

time of day, and solar activity epoch (CCIR Atlas 1967). An example in Fig. 6.1 is the predicted median *MUF(0)F2* map for June 1963 at 1800 UT. The numerical method used by CCIR is substantially derived from the method developed at the ITS (Institute of Telecommunication Sciences) at the Boulder Laboratories of the U.S. Department of Commerce by W. B. Jones and R. M. Gallet after 1960.

The basic input data consisted of the median monthly hourly values of *foF2* and *M(3000)F2* obtained from ionograms of the worldwide network of ionosondes according to the internationally agreed scaling method. The map of the ionospheric vertical sounding stations operating during the International Geophysical Year in 1958 is shown in the Introduction. The first database of observations included the years from 1954 to 1958, covering only the increasing phase of the solar cycle activity.

According to the method developed by W. B. Jones and R. M. Gallett, each station data set is represented by a Fourier time series in UT while a worldwide development of Legendre functions is applied for each Fourier coefficient. The general form of the numerical map function $\Omega(\lambda, \vartheta, T)$, also used to draw a set of charts, is given by the Fourier time series:

$$\Omega\,(\lambda,\vartheta,T) = a_0\,(\lambda,\vartheta) + \sum_{j=1}^{H}\left[a_j\,(\lambda,\vartheta)\cos\,(jT) + b_j\,(\lambda,\vartheta)\sin\,(jT)\right] \tag{6.2.1}$$

where $\lambda$ is the geographic latitude ($-90° \leq \lambda \leq 90°$), $\vartheta$ is the east geographic longitude ($0° \leq \vartheta \leq 360°$), $T$ is universal time (UTC) expressed as an angle

$(-180^\circ \le T \le 180^\circ)$, $H$ is the maximum number of harmonics used to represent diurnal variation. The Fourier coefficients, $a_j(\lambda, \vartheta)$ and $b_j(\lambda, \vartheta)$, vary with the geographic coordinates, and are represented by series of the form:

$$a_j\ (\lambda, \vartheta) = \sum_{k=0}^{K} U_{2j,k} G_k\ (\lambda, \vartheta)\,, \quad j = 0,1,2, \ldots, H \tag{6.2.2}$$

$$b_j\ (\lambda, \vartheta) = \sum_{k=0}^{K} U_{2j-1,k} G_k\ (\lambda, \vartheta)\,, \quad j = 0,1,2, \ldots, H. \tag{6.2.3}$$

The exacting choice of the functions, $G_k\ (\lambda, \vartheta)$, is determined by specifying the integers $k(k_0,\ k_1,\ \ldots,\ k_i,\ \ldots, k_m = K)$ where $i$ is the order in longitude. Therefore, a numerical map can be written explicitly in the form:

$$\begin{aligned}\Omega\ (\lambda,\vartheta,T) = & \sum_{k=0}^{K} U_{0,k} G_k\ (\lambda,\vartheta) \\ & + \sum_{j=1}^{H} \left[ \cos\ (jT) \sum_{k=0}^{K} U_{2j,k} G_k\ (\lambda,\vartheta) + \sin(jT) \sum_{k=0}^{K} U_{2j-1,k} G\ (\lambda,\vartheta) \right]\end{aligned} \tag{6.2.4}$$

where $G_k\ (\lambda, \vartheta) = \sin^{q_i} X \cos\lambda^i \sin j\vartheta$ is the geographic coordinates function. The modified magnetic dip $X$ follows the relation, $X = \arctan(I/\sqrt{\cos}\,\lambda)$, on the place where $I$ is the magnetic dip. Here the integer $q_i$, $i$ = 0, 1,..., $m$, denotes the highest power of sin $X$ for the $i$-th-order harmonics in latitude and $m$ denotes the highest order in longitude. The complete table of these functions is given in Jones et al. (1969).

The coefficients $U_{2j,k}$ and $U_{2j-1,k}$ which define the function $\Omega(\lambda, \vartheta, T)$ in Eq. 6.2.4 of the numerical map of the given characteristic for the indicated month and level of solar activity, are provided for each month of the year for *foF2* and *M(3000)F2* and for two levels of solar activity: $R_{12} = 0$ and $R_{12} = 100$. For every $R_{12}$, 12-month running mean sunspot number, different from 0 or 100 the coefficients $U_{2j,k}$ may be evaluated by a linear interpolation of the previous values for $R_{12} = 0$ and $R_{12} = 100$ or by an extrapolation for $R_{12} > 100$. Linear interpolation is also recommended for intermediate times while $R_{12}$ is to be kept at the saturation value of 150 even when $R_{12}$ exceeds this upper threshold value.

The coefficients for numerical maps of the monthly median values of *foF2* and *M(3000)F2* recommended for long-term prediction, known as "Oslo" coefficients as approved during the CCIR Assembly in Oslo 1966 are presented in the CCIR Report 340. This first set of numerical coefficients describes *foF2* and *M(3000)F2* in terms of geographical latitude, longitude, modified dip latitude and UT consists of 24 maps of 988 coefficients each, one for each month of the year and for two levels of solar activity expressed in $R_{12}$.

However, W. B. Jones and R. M. Gallet demonstrated that the use of a set of functions, which adequately represent the measured ionospheric data, can lead to instabilities, unreasonable results, or even negative values over the oceans and in other areas for which data are sparse or completely lacking. To avoid these problems it was necessary to collect as many ionospheric data as possible, which only happened after the IGY when the number of the stations almost tripled. The problem was partially solved by introducing "virtual stations", called screen points, obtained by shifting the data of coastal stations onto the oceans along the same geographical latitude, thereby using the same stations twice in the analysis. Finally, evaluated data from two sets of virtual points referred to as B and C points (Fig. 6.2) was added to the observed monthly median data from real vertical sounding stations, referred to as A points.

The data from the B points were obtained by prediction, using a correlation of the available data at the same station. The method applied an interpolation of the ionospheric characteristic by a polynomial of first or second order *versus* an index of solar activity. Obviously, A and B data were available only for land based regions. Whereas C data, used as reference points, were obtained by an interpolation of the ionospheric characteristic considered as a variable of only one coordinate, i.e. latitude. The diurnal variations of B and C points were also represented as a Fourier series. This is how, overall, the data distribution from A, B, and C points contributed to determine the numerical coefficients. The method described above gave good results for that period, especially from the perspective of ionospheric forecasters. A final presentation is generated as a set of UT hourly maps for each month showing the two propagation parameters used by the radio HF frequency planners: *MUF(0)* and *MUF(4000)* (Fig. 6.3).

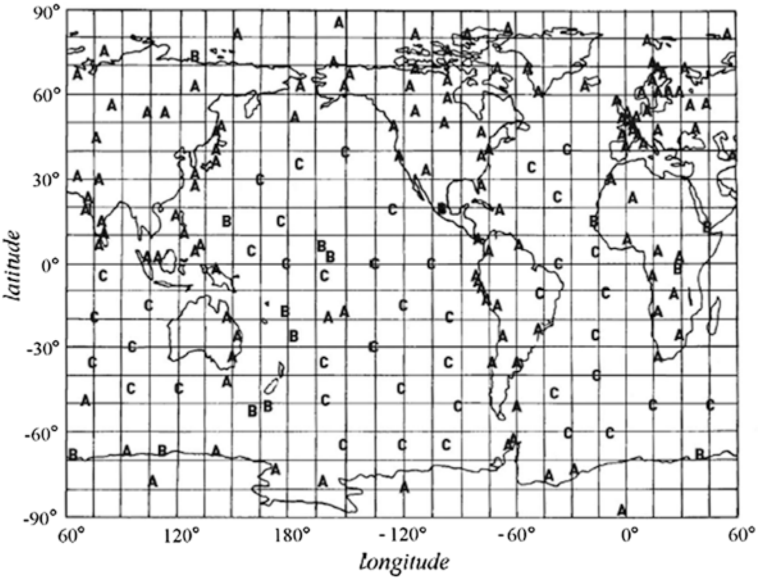


**Fig. 6.2** Map of A, B, and C type of data

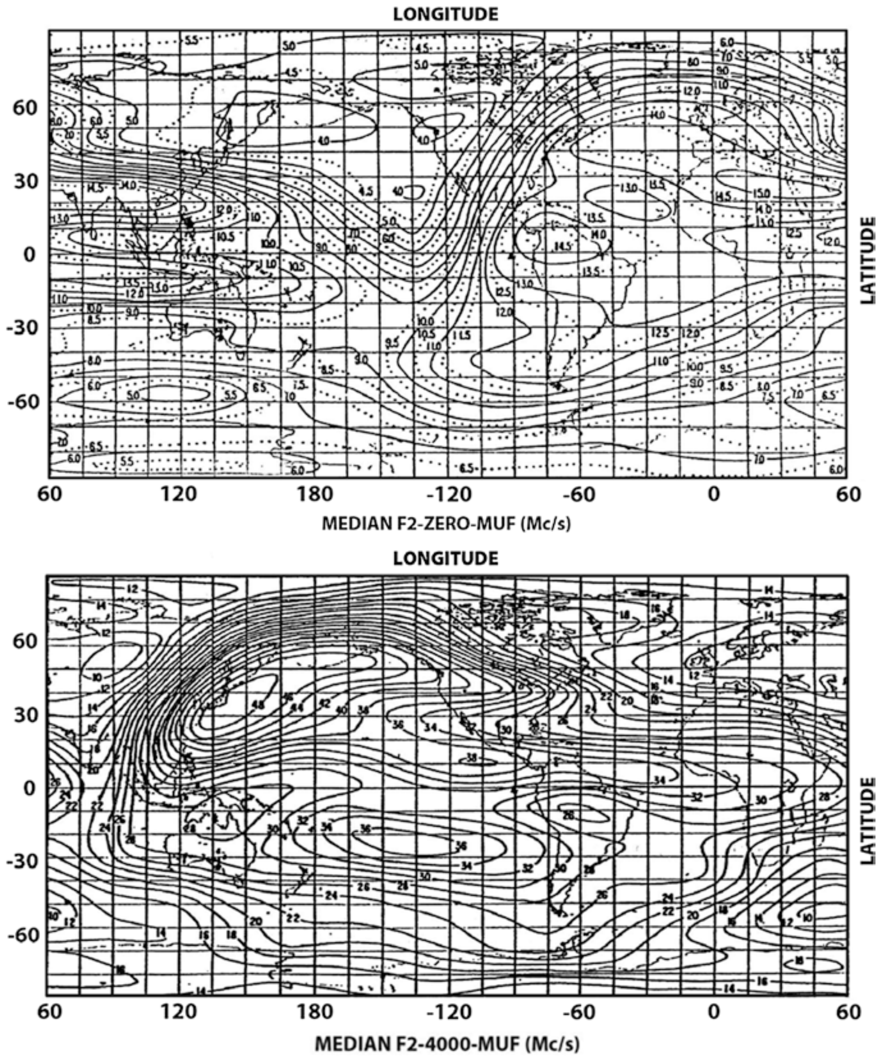


**Fig. 6.3** An example of UT hourly maps of *MUF(0)* and *MUF(4000)* in MHz

### 6.2.1.2 URSI *foF2* Prediction Maps

A very important improvement to the CCIR *foF2* global prediction model was proposed by Rush et al. (1983), and later revised by Rush et al. (1989), in order to solve the problem of when ionosonde measurements are sparse or nonexistent. They used a physical model of the ionosphere to generate values of *foF2* for mid-latitudes including those places where observations were not available. The values determined with the theoretical model were added to those from observations, before applying the spherical harmonics mapping procedure, thus contributing to generate the numerical coefficients, which produce global maps of the F2 layer critical frequency for a specific month, following the method of W. B. Jones and R. M. Gallet. The results of this method show that the description of the global behaviour of the *foF2* is consistent with the structure of the mid-latitude F2 layer and agrees with other observations such as topside soundings.

They used the time dependent ion continuity equation to calculate the theoretical values of *foF2*. If ion density $N_i$, is equal to electron density $N_e$, the values obtained by the solution of the ion continuity equation:

$$\frac{\partial N_i}{\delta t} + \nabla(N_i \mathbf{V}_i) = P_i - L_i \tag{6.2.5}$$

where $P_i$, is the ion production rate, $L_i$ the loss rate, and $\mathbf{V}_i$ the transport velocity, can be used to directly determine *foF2* as in Eq. 5.4.14.

Models of the neutral composition are used to obtain parameters such as neutral temperature production, loss and diffusion rates, neutral winds, and geomagnetic field, all of which are required for the solution of the ion continuity equation. The transport of ionospheric plasma is due principally to diffusion, neutral winds along the lines of the geomagnetic field, and electromagnetic drift perpendicular to the field lines. The effects of neutral winds and their correlation with the geomagnetic field were considered very important for the solution of the continuity equation.

The numerical coefficients, representing the global variation of *foF2* are consistent with the physical processes that govern the F region, and were arranged in the same format as in the CCIR. These new coefficients derived from observed data from the worldwide network of vertical incident ionosonde stations (A data) were then used to generate predicted data (B data), and theoretical data points (D data) obtained from the time dependent continuity equation at mid and high-latitudes. Various comparisons between the diurnal behaviour of the *foF2* obtained from the new coefficients named URSI and that obtained from the old CCIR coefficients demonstrated that the new method gives better results.

The International Reference Ionosphere, the model that gives electron density profile *versus* height, uses both the CCIR and URSI coefficients, currently know as ITU-R coefficients, to produce maps as input data of *foF2* and *M(3000)F2* over the planet when these data are not available from users or from any other assimilation method. In fact, it is recommended to use the CCIR model (Oslo coefficients) over land and the URSI model over the ocean areas and use the URSI model as the overall default model. Extensive comparisons by ground-based and in-situ measurements show significant improvement of the IRI predictions with this combined use of the two sets of coefficients. Furthermore, new studies on ionospheric mappings are currently underway by the IRI Working Group including a neural network approach and a real-time assimilative mapping technique.

#### 6.2.1.3 The Australian Method

A further improvement in ionospheric mapping was subsequently provided by Fox and McNamara (1988). They developed a method which merged the features of the Australian Ionospheric Prediction Service (IPS) maps and the maps obtained with the Jones and Gallet method modified and implemented by C. M. Rush et al. The IPS maps are drawn taking into account the importance of some physical insights. They are based on an extended bank of data and on the ionospheric index *T* that gives regression lines of *foF2 versus T* with a better correlation than other indexes. For the new world maps of *foF2* they applied a linear regression analysis of the median monthly values against the index *T* thus obtaining values of *foF2* for $T = 0$ and $T = 100$.

They provided additional data in order to fill in the gaps in the world distribution prior to mapping, by using the ITS *foF2* coefficients computed by Rush et al. valid for mid-latitudes. For low and high-latitude regions, the database was supplemented using a procedure known as "refilling", derived from IPS maps to which an interpolation is applied on geomagnetic coordinates at a given modified dip latitude.

A modified W. B. Jones and R. M. Gallet method was applied to the above mentioned data in order to produce global maps. Two important modifications were introduced to this method. The first is that with the new method the data from all the years are used to obtain maps at $T = 0$ and $T = 100$ by an easier interpolation or extrapolation. The second difference concerns the choice of the basic functions used for mapping. In reality, a good description of the behaviour of *foF2* global distribution, outside the equatorial region, is provided by means of lower-order functions than those used by Jones and Gallet, while within the equatorial region the behaviour is best described using high-order functions. These considerations are applied through two kinds of functions, the first used globally and the second only below the same cut-off latitude. Finally, a new set of updated coefficients was calculated again using data obtained theoretically over the oceans by Rush et al. (1989).

## 6.2.2 *Regional Prediction Maps*

From the previous discussion it is clear that the current global CCIR/URSI maps are based on data approximately 50 years old. They cannot take into account secular changes, important changes in the Earth's magnetic field, continuously collected *foF2* and *M(3000)F2* data, an extensive collection of data over satellite links measuring *TEC*, substantial advances in modelling and understanding of solar-terrestrial relationships, or the characteristics of the polar and equatorial ionospheric regions. Moreover, there is the particular advantage of regional study programmes for both terrestrial and trans-ionospheric applications.

A high density of ionospheric observatories over a restricted region of Europe and high quality of validated measurements were the motivating factors for the COST 238 PRIME action which proposed new methods to improve the performance of ionospheric mapping for retrospective studies, long-term prediction, and instantaneous specification. Some of these techniques were first developed for worldwide mapping and then adjusted for application in restricted regions, for example MQMF2R by A. V. Mikhailov and Yu. L. Teryokhin in 1992, while others including UNDIV by R. Leitinger in 1995, SIRM by Zolesi et al. (1993), and all Single Station Models (SSMs) were instead devised to be valid only in a limited area whatever simple or complex application they served.

The MQMF2R method was developed for the monthly median values of *foF2* and it was based on the SSM and on the multiquadric method (MQ) for spatial approximation. The multiquadric method was first developed for worldwide monthly median mapping but was subsequently adapted for regional and instantaneous mapping. An improved version of the ionospheric index *MF2* was introduced in this method to derive *foF2* regressions from solar activity levels for each ionosonde station.

The UNDIV model is a simple long-term mapping procedure based on a bilinear regression analysis in latitude and longitude of the hourly median values of *foF2* and *M(3000)F2* available over a region. It was first recommended by the COST 238 action for use in the European area for both *foF2* and *M(3000)F2* characteristics. In the COST 251 action, after tests in comparison with other developed methods, it was adopted only for *M(3000)F2*. UNDIV assumes a parabolic dependence of *foF2* on $R_{12}$ and a linear dependence of *M(3000)F2* on $R_{12}$.

#### 6.2.2.1 Simplified Ionospheric Regional Model

The SIRM (Simplified Ionospheric Regional Model) is a regional ionospheric model developed by Zolesi et al. (1993) to predict the key standard vertical incidence ionospheric characteristics such as *foF2*, *M(3000)F2, h′F, foF1,* and *foE* in a restricted area. The European area was chosen because the spatial resolution of the maximum electron density measurements was sufficiently high in comparison to the typical horizontal scale size of ionospheric dynamic phenomena. The model is based on the Fourier coefficients derived from an analysis of monthly median values of ionospheric characteristics measured at stations in the limited longitudinal range covering 40° where large variability in behaviour at mid-latitude should not be expected as a function of geographic longitude. Therefore, only model dependence by geographical latitude is taken into account.

The first step of the procedure is a linear regression analysis of the monthly median values for a given ionospheric characteristic $\Omega_{h,m}$ taken at local or universal time against the solar index $R_{12}$:

$$\Omega_{h,m} = \alpha_{h,m} R_{12} + \beta_{h,m} \tag{6.2.6}$$

The parameters $\alpha_{h,m}$ and $\beta_{h,m}$ are two matrices of $24 \times 12 = 288$ coefficients for each hour of the day, $h$, and for each month of the year, $m$. Once they are defined following the procedure explained below, *foF2* and *M(3000)F2* can be predicted for any solar epoch providing that a long-term prediction of the 12-month running mean sunspot number $R_{12}$ is available.

The second step is a Fourier analysis of the data for two fixed values of solar activity $R_{12} = 0$ and $R_{12} = 100$:

$$\Omega_{h,m} = A_0 + \sum_{n}^{l} A_n \sin(n\omega t + Y_n) \tag{6.2.7}$$

where $n$ is the harmonic number from $n = 1$ through $l = 144$, $\omega$ is the fundamental pulsation, and $t$ the time in hours. In order to reproduce the monthly behaviour more accurately the Fourier analyses were performed month by month. Considering that the 12 pairs of Fourier coefficients $A_n$ and $Y_n$ for every different month show a linear dependence on solar activity and on geographic latitude $\varphi$, this can be expressed as:

$$A_n = (a_n^1\varphi + a_n^2)R_{12} + a_n^3\varphi + a_n^4 \tag{6.2.8}$$

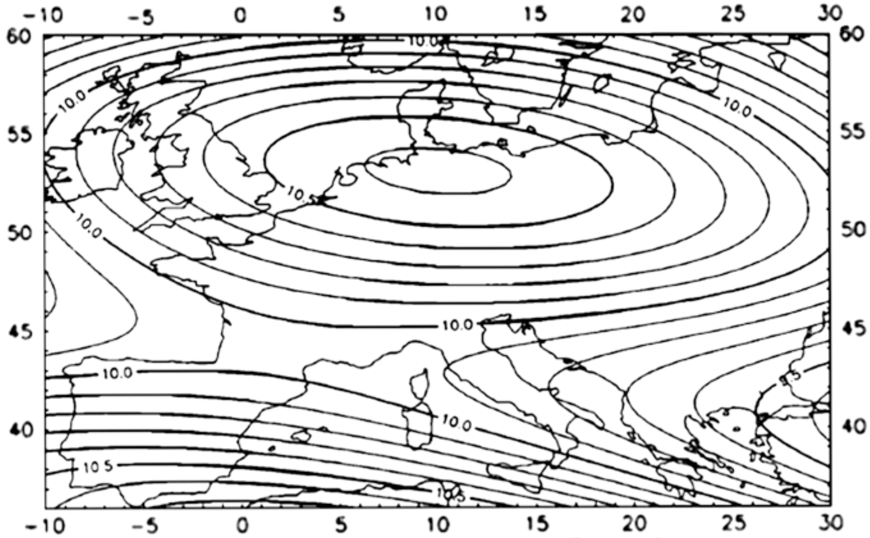
$$Y_n = (b_n^1\varphi + b_n^2)R_{12} + b_n^3\varphi + b_n^4$$

The numerical coefficients $a_n^j$ and $b_n^j$, with $j = 1, 2, 3$ and 4, can be easily calculated by a linear regression of the Fourier coefficients of every ionosonde station *versus* their latitude. Starting from a simple model for each ionosonde station produced by a linear regression analysis of the ionospheric characteristic *versus* solar activity index $R_{12}$, it is shown that 12 dominant Fourier coefficients are sufficient to reproduce the main features of the combined diurnal, seasonal, and solar cycle behaviour of the mid-latitude ionosphere under median conditions.

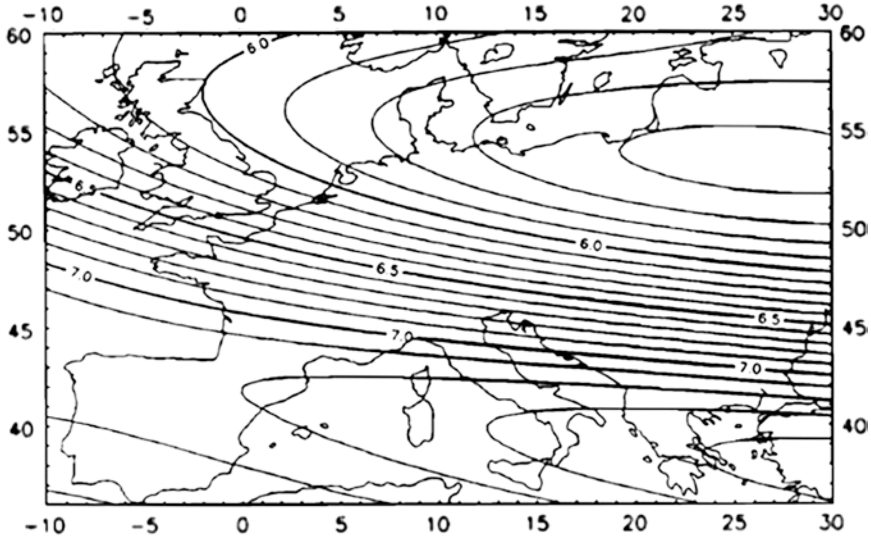
The previous set of equations describes the simplified ionospheric regional model (SIRM), the potential applicability of which to different mid-latitude areas beyond the European region was proven by Zolesi et al. (1996). The basic model was adapted and applied to different sets of inhomogeneous periods of observed data from a sparse network of ionospheric stations in mid-latitude areas: north-eastern North America, south-eastern South America, northeast Asia, and southeast Australia. Notwithstanding the simple SIRM formulation and the reduced number of numerical coefficients involved, close agreement was found between modelled and observed monthly median values of $f_0F2$ and *M(3000)F2* under different heliogeophysical conditions.

Running the SIRM model is relatively easy, not only for its simple mathematical formulation and for the reduced number of numerical coefficients, but above all for the brief source code of the program that produces maps which agree closely both with CCIR maps and with experimental observations (see Sect. 6.5). Finally, it is important to note that SIRM is easy to use and link with other software procedures such as the IRI software program.

Figures 6.4 and 6.5 show the maps of monthly median *foF2* and *M(3000)F2* distribution, respectively, obtained using the SIRM algorithm for generating the additional screen points on a latitude-longitude grid of points spaced 1° × 1°. These maps are given, with a suitable conversion, in universal time (U.T.) for two different months representing winter (January) and summer (June) ionospheric conditions in 1982 with solar activity $R_{12} = 110$.



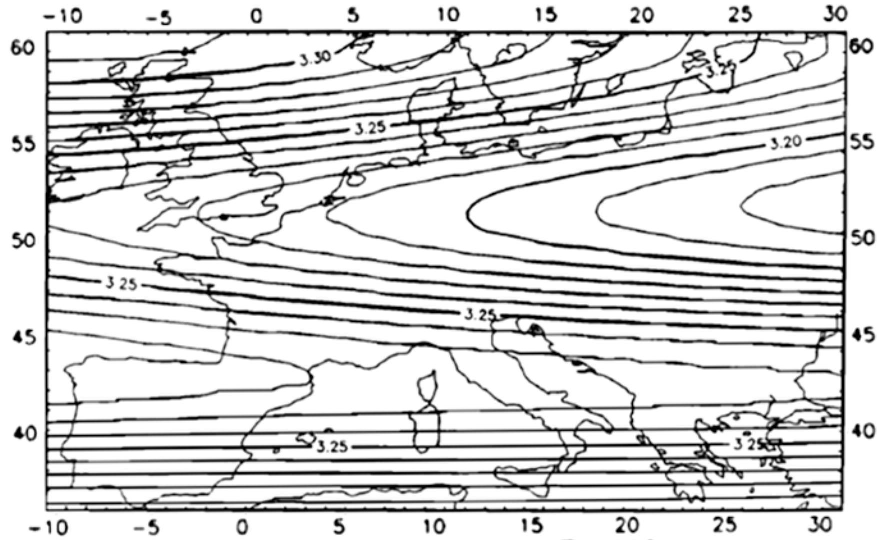
JAN. 1982 12 U.T. R=110



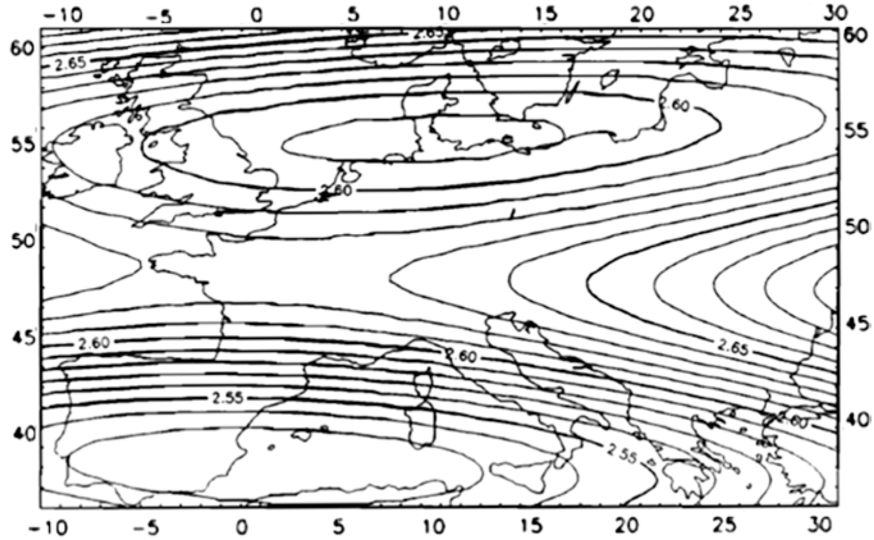
JUN. 1982 00 U.T. R=110

**Fig. 6.4** Maps of *foF2* median values calculated with SIRM for two different months with solar activity $R = 110$ (from Zolesi et al. 1993)

Ionospheric maps over the European area of monthly median *foF2* and *M(3000)F2* for different solar epochs with hourly time resolution for the long-term frequency planning services have been provided by DIAS, European DIgital upper



JAN. 1982 12 U.T. R=110



JUN. 1982 00 U.T. R=110

**Fig. 6.5** Maps of *M(3000)F2* median values calculated with SIRM for two different months with solar activity $R = 110$ (from Zolesi et al. 1993)

Atmosphere Server (DIAS) at http://www.iono.noa.gr/Dias/ since 2006 using the SIRM algorithm. DIAS is the first Europe-wide project aiming at the collection of ionospheric observations in real-time from European ionosondes and the development of added-value products for radio propagation services. Figure 6.6 gives an

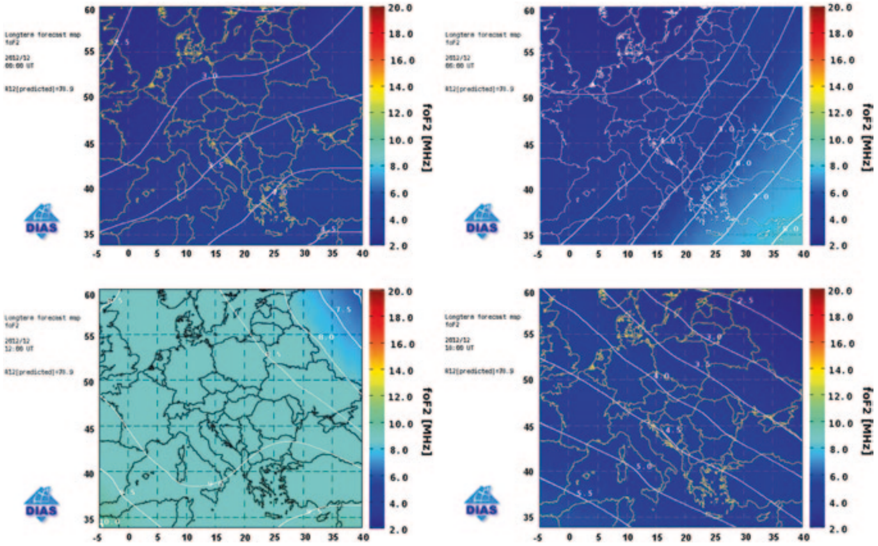


**Fig. 6.6** Sequence of hourly *foF2* long-term maps predicted by SIRM in December 2012 for DIAS (http://www.iono.noa.gr/Dias/)

example of hourly *foF2* long-term maps generated by SIRM in December 2012 for DIAS applications.

### 6.2.3 Local Single Station Models

In the past, Single Station Models (SSMs) were the typical home-made methods produced either from theory or measurement by many ionospheric prediction services both for long-term prediction and short-term forecasting. These models were very important not only for local or medium distance telecommunications purposes because of their accuracy, but also because of their ability to be interfaced with other physical models for electron density profiles and total electron content determination, or with regional and global mapping procedures. The models are obviously based on an accurate study of the *foF2* and *M(3000)F2* behaviour at a single ionosonde station with a long history of observations. Therefore, they may be considered the extreme limit of the regional models which in their immediate vicinity should be the most accurate option.

In recent decades SSMs have been widely used in other applications, for example to fill data gaps at active stations and produce new screen points for mapping ionospheric characteristics. The majority of these models apply a polynomial or even a linear regression of the monthly median values *versus* solar activity or *versus* the ionospheric index for every hour and month of the year. For example, the polynomial regression between monthly median hourly values of *foF2* and the ionospheric *S*-index given by Th.D. Xenos et al. in 1996 as:

$$foF2 = a_0 + a_1 S + a_2 S^2 \tag{6.2.9}$$

where $a_0$, $a_1$, and $a_2$ are empirical coefficients in quadratic regression. This relationship was obtained after computing the significance of the coefficients in the terms of a higher order polynomial at six mid-latitude European stations, showing that the coefficient of the cubic term is significant in only about 4–5 % of cases.

Conversely, Sole (1998) introduced a model with geomagnetic dependence based on a statistical second degree multiregression between *foF2*, the monthly ionospheric *T* index, and the monthly geomagnetic index *Ap* as follows:

$$foF2 = a + bT + cT^2 + dAp + eAp^2 \tag{6.2.10}$$

Again *a*, *b, c, d,* and *e* are empirical coefficients to be determined from ionosonde data.

Finally, SSMs are mainly used in HF communications to estimate the ionospheric condition at the midpoint of a radio link in order to evaluate the Maximum Usable Frequency rather than generate an actual map around the given ionospheric station. See for example the typical ionospheric prediction service provided by the French CNET and described by Sizun (2005).

## 6.3 Instantaneous Mapping

Ionospheric data indicative of bottom-side ionosphere peak densities and heights of the various layers collected using vertical incidence ionospheric sounder measurements are available from all over the world. However, the availability of data varies greatly in different parts of the world. Bearing in mind that data are often required for locations where no measured values are available, special interpolation is widely used for creating long-term as well as instantaneous ionospheric maps when data are collected at discrete locations (i.e. grid points). A way to achieve this is by forming a spatial (location to location) correlation model for each measured parameter and performing an interpolation. The degree to which measured values from one location influence those estimated at another depends on the distance separation weighting.

Although there are many special interpolation methods available as partly shown in Sect. 6.2, a variety of interpolation techniques have been developed to estimate the values of ionospheric characteristics at regions not covered by monitoring stations. In some of these techniques for ionospheric spatial interpolation the measured values are preserved, leading occasionally to a series of humps and troughs near the measurement locations. In other techniques the measured values are smoothed to give reasonable looking contours.

In ionospheric studies “instantaneous mapping” is defined as the technique applied when historic data-sets (retrospective mapping), simultaneously measured in real-time (nowcasting) or forecast values (forecasting mapping) of ionospheric characteristics are utilized for map generation appropriate for a single moment in time. Currently there is no internationally agreed approach to instantaneous mapping. Instantaneous maps of ionospheric characteristics, mainly used in ray-tracing connected with HF management and remote sensing systems, need to relate to different timescales for different applications. Smoothly mapping ionospheric parameters from a non-uniformly spaced set of measurement sites to a regular grid is required in many data modelling problems. The focus of interest here is with providing instantaneous electron density high profile formulations with *foF2* and *M(3000)F2* as input parameters in models described in Chap. 5.

### 6.3.1 Contouring Techniques

Instantaneous mapping methods developed over recent decades are based on computer contouring techniques, like any source of two-dimensional special representation of data, in which spatial interpolation is applied to irregularly dispersed measurements to generate corresponding values on a uniform geographical latitude-longitude grid. Various approaches to the determination of optimum weighting factors exist. Some of these exactly reproduce the original measurement values; others smooth them out to give less marked gradient maps. The contouring techniques used in recent years, involve the generation of a uniform latitude-longitude grid of values by minimum curvature, inverse-distance, or Kriging procedures.

Critical testing of three interpolation methods against measured data conducted during the PRIME (Prediction and Retrospective Ionospheic Modelling over Europe) COST238 action (1991–1995) led to the conclusion that the Kriging approach could be the best one for ionospheric studies and applications. Kriging is a well-known geostatistical interpolation technique firstly introduced in mining engineering and named after its originator D.G. Krige.

This technique is based on the estimated spatial covariance structure of observed data, i.e. two-dimensional weighted averages, where the weights are a function of the separation distance from the semivariogram. It produces a minimum unbiased estimate and provides estimates at nearby locations as shown by Oliver and Webster (1990). Kriging provides original measured values and the changing correlation distances on separate occasions and over different areas at the same time are taken into account. However, as it behaves nonlinearly with respect to spatial coordinates, it is clear that limited data sets or not validated data might produce non-realistic functions (see Chap. 8).

Nevertheless, the Kriging method is successfully used to spatially interpolate derived *foF2* and *MUF(3000)F2* characteristics as defined in Sect. 3.3.4 from a sparse network of ionospheric sounder observations. Examples of measurement maps of *foF2* and *MUF(3000)F2* are given in Figs. 6.7 and 6.8 for 28 March 2000, at 1200 UT. The crosses in these figures indicate measured values at ionosonde stations, representing different ionospheric conditions over an area of −10°E to

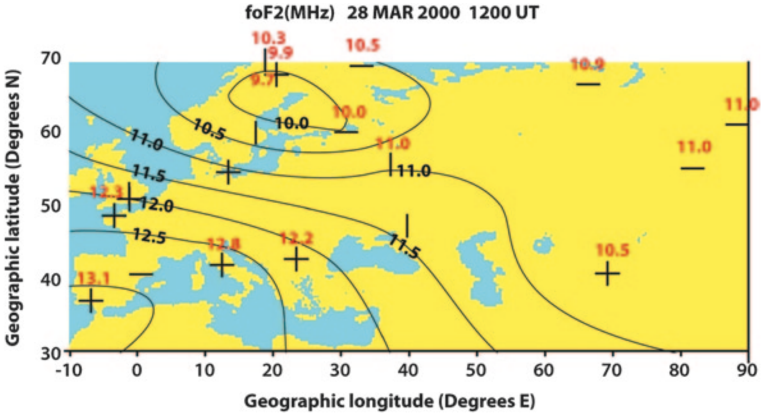


**Fig. 6.7** Map derived from measured values of *foF2* for 28 March 2000 at 1200 UT

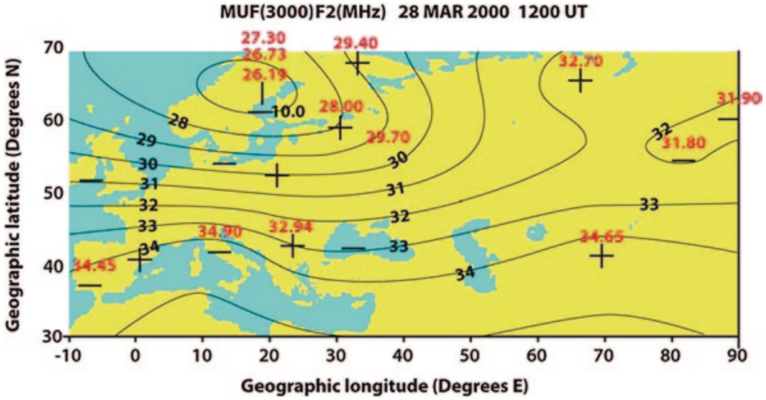


**Fig. 6.8** Map derived from measured values of *MUF(3000)F2* for 28 March 2000 at 1200 UT

90°E in longitude and 30–70°N in latitude. In these examples grid resolution is 2.5° in latitude and 5° in longitude produced using Kriging as implemented in the commercial Surfer software.

### 6.3.2 *Instantaneous Mapping Based on Additional Screen-Point Values*

A much more complex approach involves artificial screen-point values in regions remote from measurement locations to constrain the contours within “acceptable” bounds. Generally the number of measurement locations is insufficient for the production of completely accurate maps, even over a restricted geographical region. The methods involve choosing the form of smoothing functions between

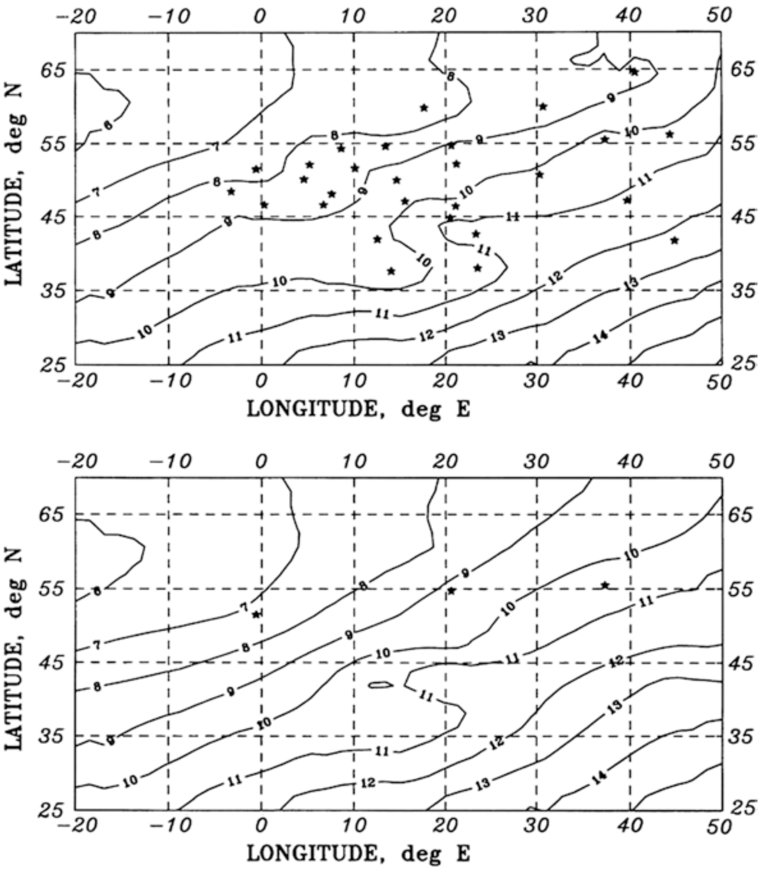


**Fig. 6.9** Maps of *foF2* determined by the MQMF2-IM technique on 15 April 1980 at *0800 UT*. *Top panel* from 28 sets of measurements; *lower panel* from 3 sets of measurements (from Bradley 1995)

measurements and of adopting screen-point values in regions where no measurements are available. Screen-point values may be taken from monthly median predictions using real or effective sunspot numbers derived by optimizing the prediction fit at the measurement locations. Alternatively, they can be derived from empirical formulations for locations where measurements were taken in the past.

The PRIME recommended instantaneous mapping method is the MQMF2-IM (Multiquadric method with new ionospheric index *MF2* for instantaneous mapping) developed by Mikhailov and Mikhailov (1995) as an extension of the technique applied by the same co-authors in long-term mapping, as mentioned in Sect. 6.2.2. The MQMF2-IM for optimum performance requires a minimum of 5 observational values for *foF2* and 3 observational values for *M(3000)F2* with additional screen-points generated by a Single Station Model. Figure 6.9 presents *foF2* maps obtained using the MQMF2-IM method when different numbers of measurement points are taken into account on 15 April 1980 at 0800 UT, top panel with observed data points only, and lower panel with modelled results. The effect of the screen points in absence of the measured values is evident in preserving geographical structure so that two regional *foF2* maps are quite similar.

## 6.4 Nowcasting: Real-Time Ionospheric Specification

The possibility of merging ionospheric models with immediate observations is the "*conditio sine qua non*" for a nowcasting service, providing users nowcasting maps as well as an accurate representation of current ionospheric conditions. However, potential errors in the autoscaling procedure (see details in Chap. 3) or the impossibility of automatically scaling difficult ionograms, often occurring during geomagnetic disturbed periods or due to peculiar sporadic E layer conditions, may introduce errors in parts of the map.

Combining models and real-time measurements to generate a coherent global or regional picture, consistent with the most recent observations, is the basis for a nowcasting service. The number of such services has grown significantly in recent years, but attention is concentrated below on the IPS global nowcasting maps, and two regional approaches for real-time ionospheric specification over limited areas of Europe and Australia.

### *6.4.1 IPS Global and Regional Nowcasting Maps*

The IPS applies the T-INDEX method to generate predictions of the ionospheric *foF2* and *M(3000)F2* characteristics for any time and location in the world. The $T$ index is an ionospheric index derived from observed *foF2* and sunspot number $R_{12}$ data. IPS provides world ionospheric *foF2* maps, which are updated hourly by the automated interpretation of ionograms from around the world. The *foF2* values in the mapping process are based on the median of the last 2 h of station data (4 values per hour for a ionosonde operating at 15 min intervals) to avoid autoscaling problems.

The mapping technique is based on the world monthly median maps of *foF2*, assumed to represent the "quiet ionosphere", available for each UT hour for 12 months and two levels of the ionospheric $T$ index, constructed from the long history of worldwide ionospheric measurements. Using observed autoscaled *foF2* values from the contributing stations and applying an interpolation of the quiet maps, it is possible to generate a new and effective $T$ index for each location. The new irregular grid of individual $T$ index is then converted into a regular grid by a two dimensional interpolation. This new effective index set is finally used to define the regular grid of the *foF2* values and draw the map. The technique is applied for both global and regional mapping as shown in Fig. 6.10.

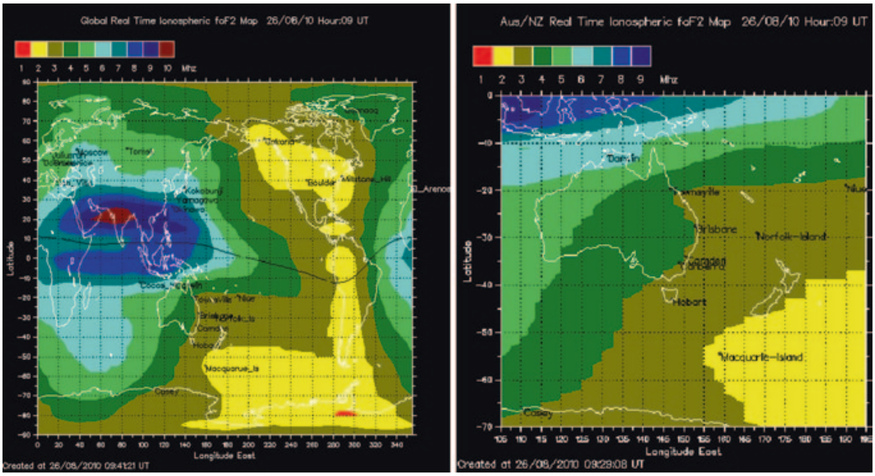


**Fig. 6.10** Global and regional real-time ionospheric *foF2* map for 26 August 2010 at 0900 UT (http://www.ips.gov.au/)

### 6.4.2 SIRMUP

DIAS ionospheric maps over the European area of real-time *foF2* and *M(3000) F2* for individual epochs in 15 min time resolution for nowcasting frequency management have been regularly provided since 2006 following the SIRMUP method. The Simplified Ionospheric Regional Model Updated in real-time (SIRMUP) is a nowcasting procedure to provide regional maps of the key ionospheric *foF2* and *M(3000)F2* characteristics over Europe developed by Zolesi et al. (2004). It is based on the Simplified Ionospheric Regional Model (SIRM), a long-term prediction model described in Sect. 6.2.2.1, updated with real-time ionospheric observations to produce nowcasting maps. Currently modern digital ionosondes are employed worldwide as a tool to remotely sense the ionosphere and identify its structure and properties. Most digital ionospheric stations have the capability of automatically scaling and transmitting in real-time all important parameters characterizing the state of the ionosphere and the propagation of radio waves.

As such, they can be employed for real-time specification of the operational parameters of the telecommunications system in restricted areas worldwide. The SIRMUP operates by a rapid conversion of real-time autoscaled *foF2* and *M(3000)F2* data from ionosondes into the driving parameter of the SIRM, which is an effective sunspot number $R_{eff}$, calculated by a method introduced by Houminer et al. (1993). With SIRMUP, the $R_{eff}$ is chosen to give the best fit between the SIRM model output and the actual measurements obtained from ionosondes located in the mapping area.

The main steps of the SIRMUP regional nowcasting mapping technique can be summarized as follows: to initiate the procedure, a starting value of $\Omega$ ($\Omega$ stands for the *foF2* or the *M(3000)F2*) is calculated using the SIRM model for a given

$R_{12}$; then an iterative procedure is applied to adjust the sunspot number used by SIRM until the mean square error between the SIRM calculation and real observations is minimized. The sunspot number, giving the minimum mean square error, is called the effective sunspot number $R_{eff}$. The mean square error is given by:

$$\Delta = \frac{1}{n}\sum_{i=1}^{n}(\Omega_{obsi} - \Omega_{calci})^2 \tag{6.2.11}$$

where $n$ is the number of ionospheric stations available, $\Omega_{\text{obsi}}$ is the observed value of the $\Omega$ characteristic at the station $i$, and $\Omega_{\text{calci}}$ is the corresponding SIRM calculated value at station $i$. When $R_{eff}$ is known, the new SIRMUP grid of *foF2* and *M(3000)F2* based on SIRM calculation can be generated.

The main input parameters of the SIRMUP nowcasting procedure are: (a) real-time *foF2* and *M(3000)F2* values from ionosondes autoscaled at hourly or 15 min intervals; and (b) grids of the median monthly *foF2* and *M(3000)F2* values from SIRM obtained using predicted values of the sunspot number $R_{12}$ for the specific month. Output values are: (a) nowcast grid values for *foF2* and *M(3000)F2* presented as numerical tables spaced $1° \times 1°$ for all 24 h and possibly every 15 min; and (b) routine graphical procedures to represent the isolines as maps. It should be noted that computing time to generate the grid values is less than a few seconds.

SIRMUP was developed and applied to several inhomogeneous ionospheric data sets in different European mid-latitude areas. The agreement between the SIRMUP real-time values and observed data of *foF2* and *M(3000)F2* was demonstrated to be satisfactory. The SIRMUP simulation results, in terms of both *foF2* and *M(3000)F2* ionospheric characteristics regarding various quiet and disturbed ionospheric conditions, were shown by Pietrella et al. (2009) to achieve a high degree of efficiency in real-time mapping of propagation conditions over Europe.

The final output from the SIRMUP nowcasting procedure are maps of *foF2* and *M(3000)F2* covering the European area from −5° E to 40° E in longitude and 34° N to 60° N in latitude in the case of the DIAS project application (http://www.iono.noa.gr/DIAS/). SIRMUP has also been applied locally to provide ionospheric nowcasting maps of *foF2* and *M(3000)F2* for a region of interest to a specific user, like for example the Central Mediterranean (Fig. 6.11).

Both IPS and SIRMUP nowcasting procedures are based on the concept that a long-term prediction map provides a background to be updated with real-time observations from ionosonde stations. The differences consist in the use of diverse effective indices: ionospheric interpolated $T$ index by IPS, and effective sunspot number $R_{eff}$ by SIRMUP.

Regional SIRMUP and global IRI ionospheric models can provide a valid tool for obtaining a real-time three-dimensional (3-D) electron density map of the ionosphere. Figure 6.12 shows an example of a 3-D model where maps of the electron density (electrons/cm$^3$) at the fixed height 210 km are obtained using the IRI model with the *foF2* and *M(3000)F2* input parameters calculated with the URSI coefficients on the left, and IRI with the input parameters calculated with the regional nowcasting method SIRMUP on the right (also see related Fig. 5.10). These maps are valid for three different smoothing coefficient values (sigma).

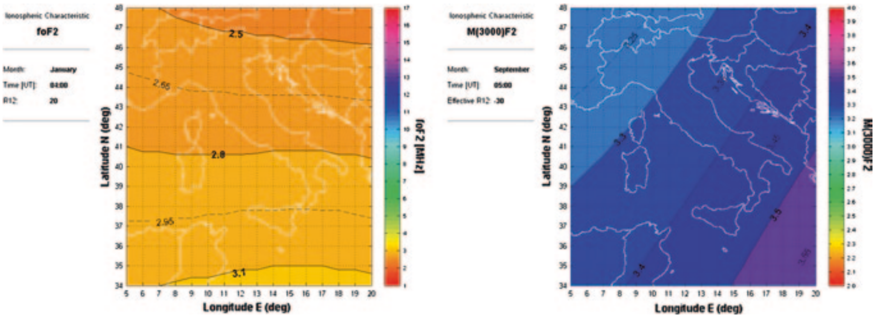


**Fig. 6.11** Examples of local nowcasting maps for *foF2* and *M(3000)F2* over the Central Mediterranean (http://www.ingv.it/)

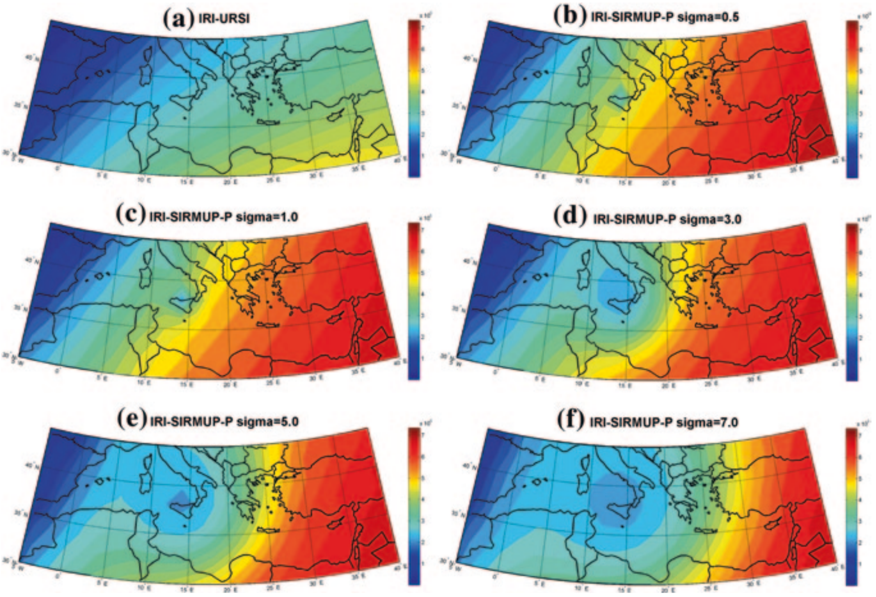


**Fig. 6.12** Maps of electron density (electrons/cm$^3$) at the fixed height 210 km obtained by (**a**) IRI-URSI, and IRI-SIRMUP-P for different smoothing coefficient values (from Pezzopane et al. 2011)

## 6.5 Testing Procedures

In the past, testing of ionospheric mapping methods was conducted directly by the authors, or carried out for particular cases and there was no internationally agreed procedure. Only during the ionospheric COST actions were several mapping methods tested against a large database of measurements not used in map generation.

It should be noted that in any testing, quality checks on the testing databases result in the elimination of most of the unreliable data. For long-term mapping, months with gaps of total length greater than 15 days, or single gaps of length greater than 10 days, or extreme values exceeding 20 % of the total number of valid values for that month should be eliminated. Hours with a number of valid values less than 5 should also be eliminated. For instantaneous mapping, extreme values are always eliminated. A threshold probability of 0.99 should be used to define the extreme values for each ionospheric station.

Comparisons of measurements and predictions for the testing stations could be carried out, using all the valid monthly median data available at the test stations. The root mean square (*RMS)* error is given by:

$$\sigma = \left( \frac{1}{N} \sum_{i=1}^{N} (p_i - m_i)^2 \right)^{1/2} \tag{6.2.12}$$

where $N$ is the total number of samples, $p_i$ is the value predicted by the model for sample $i$, and $m_i$ is the measured value for sample $i$.

For example, the reliability of the SIRM mapping technique was tested following this procedure and the results show that the overall *RMS* error from SIRM was slightly smaller than the *RMS* error from the recommended but more time-consuming ITU model. This validation test proves that SIRM performance is satisfactory for describing median ionospheric conditions in a restricted area of the mid-latitudes.

## Suggested Readings

Anderson DN (1973) A theoretical study of the ionosphere F region equatorial anomaly. Planet Space Sci 21:409–419

Belehaki A, Cander LjR, Zolesi B, Bremer J, Juren C, Stanislawska I, Dialetis D, Hatzopoulos M (2005) DIAS project: The establishment of a European Digital Upper Atmosphere Server. J Atmos Sol-Terr Phys 67:1092–1099

Bilitza D, Sheikh NM, Eyfrig R (1979) A global model for the height of the F2-peak using M(3000)F2 values from the CCIR numerical map. Telecommun J 46:549–553

Bilitza D, Rawer K, Pallaschke S, Rush CM, Matuura N, Hoegy WR (1987) Progress in modeling the ionospheric peak and topside electron density. Adv Space Res 7:5–12

Bradley PA (1995) PRIME (Prediction and Retrospective Ionospheric Modelling over Europe), COST Action 238 Final Report. Commission of the European Communities, Brussels

Cander LjR (2003) Towards forecasting and mapping ionospheric space weather under the COST actions. Adv Space Res 31:957–964

CCIR Atlas of Ionospheric Characteristics (1967) Comité Consultatif International des Radiocommunications, Report 340-4, International Telecommunications Union, Geneva

Davis JC (1986) Statistics and data analysis in geology. Wiley, New York

De Santis A, De Franceschi G, Zolesi B, Cander LjR (1992) Regional modelling and mapping of the ionospheric characteristic parameters by spherical cap harmonic expansion. Adv Space Res 12:279–282

Dominici P, Zolesi B (1987) A model for normal ionosphere over Rome. Il Nuovo Cimento 10C:191–208

Fox MW, McNamara LF (1988) Improved worldwide maps of monthly median foF2. J Atmos Terr Phys 50:1077–1086

Hanbaba R (1991) Les previsions de propagation ionospherique des ondes radioelectriques. Atti del II Convegno di Geomagnetismo e Aeronomia, Istituto Nazionale di Geofisica, Roma, 18–20 Ottobre 1988,109–143

Hanbaba R (1999) Improved Quality of Service in Ionospheric Telecommunication Systems Planning and Operation, COST 251 Final Report. Space Research Centre Printing Office, Warsaw

Houminer Z, Bennett JA, Dyson PL (1993) Real-time ionospheric model updating. J Elect Electron Eng 13:99–104

ITU-R (1994) HF Propagation Prediction Method. Recommendation ITU-R, P. 533, International Telecommunication Union, Geneva

Jones WB, Gallet RM (1960) Ionospheric mapping by numerical methods. Telecommun J 12:260–264

Jones WB, Gallet RM (1962) The representation of diurnal and geographic variations of ionospheric data by numerical methods. Telecommun J 29:129–147

Jones WB, Gallet RM (1965) The representation of diurnal and geographic variations of ionospheric data by numerical methods. Telecommun J 32:18–28

Jones WB, Graham R, Leftin M (1969) Advances in ionospheric mapping by numerical methods. ESSA Technical Report, ERL 107—ITS 75. Institute for Telecommunications Sciences, Boulder, Colorado, US Government Printing Office, Washington, DC

Mikhailov AV, Mikhailov VV (1995) A new ionospheric index MF2. Adv Space Res 12:93–97

Oliver MA, Webster R (1990) Kriging: a method of interpolation for geographical information systems. Int J Geog Inf Syst 4:313–332

Pezzopane M, Pietrella M, Pignatelli A, Zolesi B, Cander LjR (2011) Assimilation of autoscaled data and regional and local ionospheric models as input sources for real-time 3-D International Reference Ionosphere modeling. Radio Sci. doi:10.1029/2011RS004697

Piggott WR, Rawer K (1972) URSI Handbook of Ionogram Interpretation and Reduction. Report UAG-23, World Data Center A for Solar-Terrestrial Physics, NOAA, Environmental Data Service, Asheville

Pietrella M, Perrone L, Fontana G, Romano V, Malagnini A, Tutone G, Zolesi B, Cander LJR, Belehaki A, Tsagouri I, Kouris SS, Vallianatos F, Makris JP, Angling MJ (2009) Oblique-incidence ionospheric soundings over Central Europe and their application for testing now casting and long term prediction models. Adv Space Res. doi:10.1016/j.asr.2008.09

Rush CM, PoKempner M, Anderson DN, Stewart FG, Perry J (1983) Improving ionospheric maps using theoretically derived values of foF2. Radio Sci 18:95–107

Rush CM, PoKempner M, Anderson DN, Perry J, Stewart FG, Reasoner R (1984) Maps of foF2 derived from observations and theoretical data. Radio Sci 19:1083–1097

Rush CM, Fox M, Bilitza D, Davies K, McNamara LF, Stewart F, PoKempner M (1989) Ionospheric mapping: an update of foF2 coefficients. Telecommun J 56:179–182

Rawer K (1975) The historical development of forecasting methods for ionospheric propagation of HF waves. Radio Sci 10:669–679

Rawer K (1987) Actual problems with ionospheric mapping. Indian J Radio Space Phys 16:373–379

Samardjiev T, Bradley PA, Cander LjR, Dick MI (1993) Ionospheric mapping by computer contouring techniques. Electron Lett 29:1794–1795

Singer W, Dvinskikh I (1991) Comparison of empirical models of ionospheric characteristics developed by means of different mapping methods. Adv Space Res 11:3–6

Sizun H (2005) Radio Wave Propagation for Telecommunication Applications. Springer, Heidelberg

Solé JG (1998) Relations between hourly monthly median values of foF2 and some geophysical indices. Their application to an ionospheric single station model. Acta Geophys Pol XLVI:77–88

Stanislawska I, Juchnikowski G, Cander LjR (1996) Kriging method of ionospheric parameter foF2 instantaneous mapping. Ann Geofis XXXIX:845–852

Surfer for Windows Ver 6, Golden Software Ltd, 809 14th St, Golden, CO 80401, USA

Tsagouri I, Zolesi B, Belehaki A, Cander LjR (2005) Evaluation of the performance of the real-time updated simplified ionospheric regional model for the European area. J Atmos Sol-Terr Phys 67:1137–1146

Turner JF (1968) The development of the ionospheric index T. IPS Series_R report IPS-R11

Wilkinson PJ (1986) A review of daily ionospheric indices: current status and recent innovations. Series Title: IPS Radio and Space Services Technical Report; IPS-TR-86-02, Darlinghurst, N.S.W

Wilkinson PJ (1995) Predictability of ionospheric variations for quiet and disturbed conditions. J Atmos Terr Phys 57:1469–1481

Zolesi B, Cander LjR, De Franceschi G (1993) Simplified ionospheric regional model for telecommunication applications. Radio Sci 28:603–612

Zolesi B, Cander LjR, De Franceschi G (1996) On the potential applicability of the simplified ionospheric regional model to different mid-latitude areas. Radio Sci 31:547–552

Zolesi B, Belehaki A, Tsagouri I, Cander LjR (2004) Real-time updating of the simplified ionospheric regional model for operational applications. Radio Sci. doi:10.1029/2003RS002936

Zolesi B, Cander LjR (2004) COST 271 Action—Effects of the upper atmosphere on terrestrial and Earth-space communications: introduction. Ann Geophys 47:915–925

# Chapter 7
# Total Electron Content Modelling and Mapping

## 7.1 Introduction

Although the main purpose of the Global Positioning System (GPS) was to provide accurate positional location at any point on or above the Earth's surface at all times, for defence and civilian purposes, GPS also had a major impact on ionospheric science during the last decade of the twentieth century. Sections 3.5.2 and 4.3 described how the measurements of the existing Global Navigation Satellite Systems (GNSS), with unprecedented rates of temporal and spatial sampling on a global scale, could systematically provide values for the ionospheric total electron content (*TEC*) using dual-frequency code pseudorange and carrier phase data. These values have significantly contributed to efforts in empirical modelling and space weather event analysis, for the generation of new accurate global, regional and local ionospheric maps of vertical total electron content (*VTEC* maps), and will serve for future improvements based on the needs of navigation and other service users.

Excellent reference literature is available, introducing not only the fundamentals of GNSS and GNSS data processing, but also the significant enhancement of GNSS due to an improved understanding of ionospheric conditions (see for example Hernández-Pajares et al. 2011). Consequently, this Chapter only briefly describes: (1) one of the many *TEC* evaluation techniques from RINEX files currently in use (Sect. 7.2); (2) a method for empirically modelling total electron content in relation to its simple application in long-term ionospheric prediction (Sect. 7.3); (3) GPS *VTEC* mapping based on data from the International GNSS Service (IGS) network as the standard means for ionospheric real-time monitoring and nowcasting (Sect. 7.4); (4) regional *TEC* mapping (Sect. 7.5); and (5) some aspects of two or three dimensional ionospheric tomography using GPS (Sect. 7.6).

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_7,

## 7.2 *TEC* Evaluation Technique from RINEX Files

As described in Sect. 3.5.2, the delay produced when GNSS signals pass through the Earth's ionosphere is inversely proportional to the square of the carrier frequency, and directly proportional to the total electron content. Accordingly, various techniques using dual frequency GNSS measurements have been developed for extracting *TEC* values from the amplitude and phase of GNSS signals. Although the accuracy of the calibrated *TEC* values varies from one evaluation technique to another, in general the error effects are confined to a few TECU at mid-latitudes.

Ciraolo's technique, introduced in 1993 and continuously improved since then, starts with each RINEX file containing phase and group propagation delays at two frequencies in the L-band at $f1 = 1{,}575.42$ MHz (L1) and $f2 = 1{,}227.60$ MHz (L2), observed at a single station during a UTC day. By appropriately differencing these delays, respectively of phase and group, two quantities proportional to slant *TEC* are obtained, commonly known as "phase" and "group" (or "code") slants. Slants are affected by hardware biases (group) and offsets (phase), which are both unknowns.

The first step before using these for ionospheric investigation is an evaluation of biases and offsets, the process of "calibration" or "de-biasing". Either phase or code slants can be used for this step, but generally a combination of both is used. The resulting "levelled" slants are free of the ambiguities inherent to phase observations, and to some extent free of the multi-paths that strongly influence code slants. Taking $\Omega$ as the phase offset, $\beta$ and $\gamma$, the satellite and station hardware biases, $m$ the multi-path, and <> a weighted average operator, the three quantities obtained up to this point can be expressed as follows:

$$\text{Phase slants} \qquad S_\Phi = TEC + \Omega \tag{7.2.1}$$

$$\text{Code slants} \qquad S_G = TEC + \beta + m \tag{7.2.2}$$

$$\text{Levelled slants} \qquad S = S_\Phi - \langle S_\Phi - S_G \rangle = TEC + \beta + \gamma + \langle m \rangle \tag{7.2.3}$$

The observables used for calibration are always the levelled slants. As a single station calibration is performed, $\beta$ and $\gamma$ cannot be separated, and $\beta$ will generally indicate their sum, while the $<m>$ contribution is ignored. Under these assumptions, from a given station, the set of slants to each satellite $i$ at time $t$ is:

$$S_{it} = TEC_{it} + \beta_i \tag{7.2.4}$$

According to the thin shell approach, the slant is expressed as:

$$TEC = VTEC \sec \chi \tag{7.2.5}$$

Here *VTEC* is assumed to be the vertical *TEC*, a function of the (horizontal) position of the ionospheric (or pierce) point P identified as the intersection of the thin shell with the ray, $\chi$ being the angle between the ray itself and the vertical in

P (also see Fig. 3.20). Subsequently, *VTEC* is expressed as a proper expansion of horizontal coordinates *l, f* with one set of coefficients $c_n p_n$ at each time:

$$VTEC_{Eq}\,(l,\,f) = \sum\nolimits_n c_n p_n\,(l,\,f) \tag{7.2.6}$$

The observation Eqs. 7.2.4 will therefore be written as:

$$S_{it} = \sum\nolimits_n c_{tn} p_n\,(l_{it},\,f_{it})\sec\chi_{it} + \beta_i \tag{7.2.7}$$

The equations for the above set of slants are written for every UTC day for each station. The reference height for the computation of $\chi_{\boldsymbol{REF}}$ is set at 400 km. The horizontal coordinates are local time and K. Rawer modified dip latitude (modip). The set of equations is linear in the unknown coefficients $c_{tn}$ and biases $\beta_i$. The solution provides the actual *TEC*'s (or calibrated slants) as $TEC_{it} = S_{it} - \beta_i$, and the vertical *TEC* over the station from the expansion Eq. 7.2.6, using for *(l, f)* the coordinates of the station itself.

L. Ciraolo's technique has been used in numerous applications including the ionospheric correction information provided by Satellite Based Augmentation Systems (SBAS), for example the European Geostationary Navigation Overlay Service (EGNOS), as well as in all the *VTEC* plots throughout this book. Its latest upgrading is related to the usage of the individual arc offsets as unknowns instead of hardware biases in a single-station, multi-day solution aiming to improve processing data from low-latitude stations.

## 7.3 Total Electron Content Modelling

Total electron content modelling involves a set of procedures to establish a mathematical representation of global *TEC*, together with its temporal variations. Total vertical electron content (*VTEC*) varies with the solar cycle, month to month, and day to day. The dominant feature is the approximately 11-year cycle in which sunspot numbers vary from around 0 at absolute solar minimum to around 200 at solar maximum (see Chap. 2).

According to Davies and Smith (2000) the daytime 12-month smoothed *TEC*, $TEC_D(R_{12})$ varies approximately linearly with the smoothed sunspot number $R_{12}$ according to the equation:

$$TEC_D\,(R_{12}) = TEC_D\,(0)\,[1 + 0.02\,R_{12}] \tag{7.3.1}$$

The corresponding night-time relationship is:

$$TEC_N\,(R_{12}) = TEC_N\,(0)\,[1 + 0.01\,R_{12}] \tag{7.3.2}$$

where $TEC_D(0)$ and $TEC_N(0)$ are the day and night total vertical electron content in TECU respectively for sunspot number zero.

A comparison of the observed *VTEC* monthly median diurnal variations at the GPS ground station at HERS (50.9 N, 0.3 E) in Hailsham, with *VTEC* calculated by Eqs. 7.3.1 and 7.3.2 during low solar activity in ionospheric winter, summer, and equinox conditions, is shown in Fig. 7.1.

By using the simple empirical Eqs. 7.3.1 and 7.3.2 in a slightly modified form:

$$VTEC_D\,(R_{12}) = VTEC_D\,(0)\,[1 + A_D\,R_{12}] \tag{7.3.3}$$

$$VTEC_N\,(R_{12}) = VTEC_N\,(0)\,[1 + A_N\,R_{12}] \tag{7.3.4}$$

where $A_D$ and $A_N$ are statistically based daytime and night-time coefficients respectively, a climatological single location model for monthly median *VTEC* was

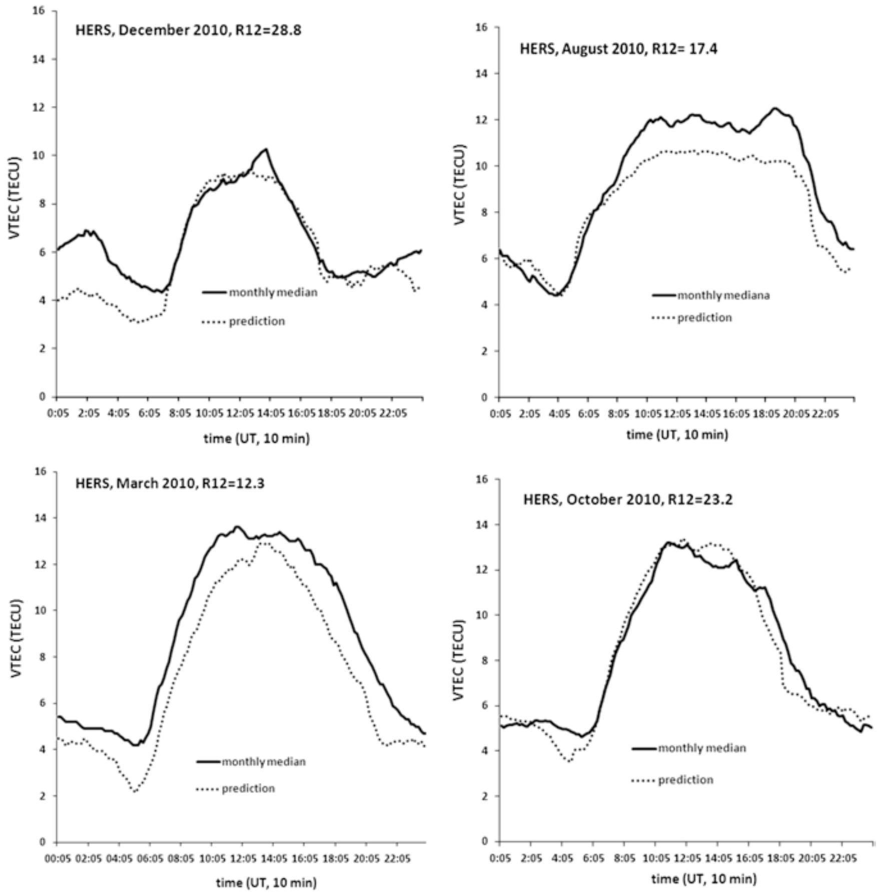


**Fig. 7.1** Comparisons of the observed *VTEC* monthly median at HERS (50.9 N, 0.3 E) in 2010 for December (*top left*), August (*top right*), March (*bottom right*), and October (*bottom left*) with the corresponding predictions

developed. This was tested during high solar activity conditions in 2000 at the high mid-latitude GPS station HERS, and the low mid-latitude GPS station NICO (35.1 N, 33.4 E) in Nicosia, Cyprus.

The estimated $A_D$ and $A_N$ coefficient values are labelled in the figures. Even though the model values tended to under-estimate *VTEC* by a few TECU during daytime at HERS (Fig. 7.2), the overall agreement between the observed and predicted monthly medians is very good. Figure 7.3 shows that this is particularly convincing for the NICO *VTEC* monthly median data.

These results demonstrate that *VTEC* ionospheric models are now quite mature, which becomes even more obvious for other much more sophisticated models, like those recently reviewed by Jakowski et al. (2011). Among these is NeQuick 2, currently very widely used for both research and training purposes. This is, the

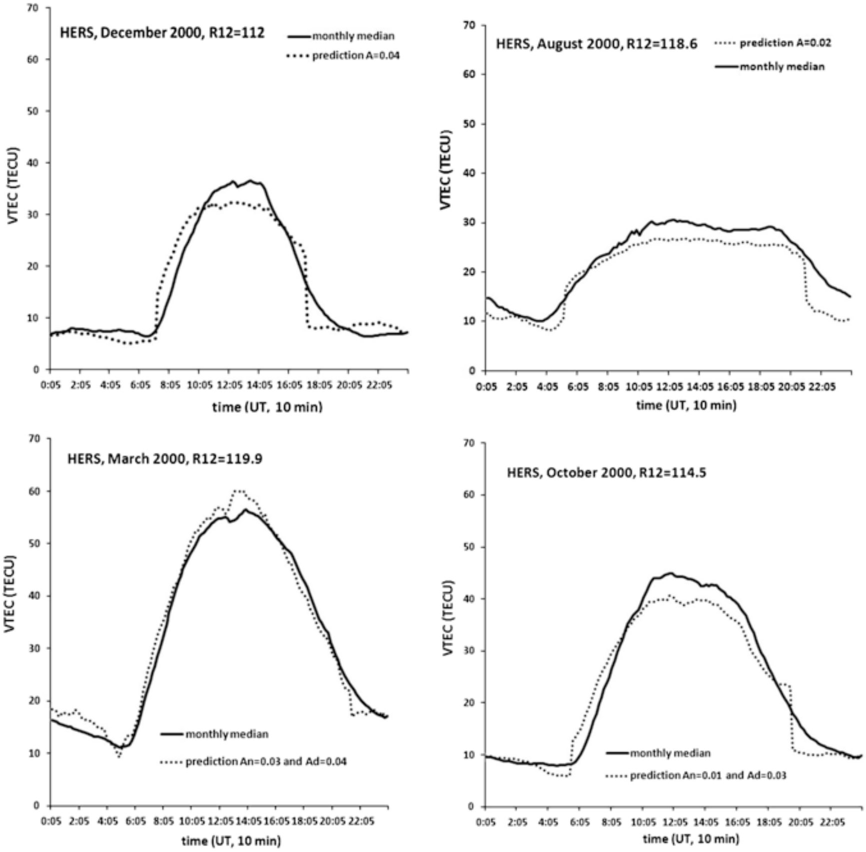


**Fig. 7.2** Comparisons of the observed *VTEC* monthly median at HERS (50.9 N, 0.3 E) in 2000 for December (*top left),* August (*top right*), March (*bottom right*), and October (*bottom left*) with the corresponding predictions

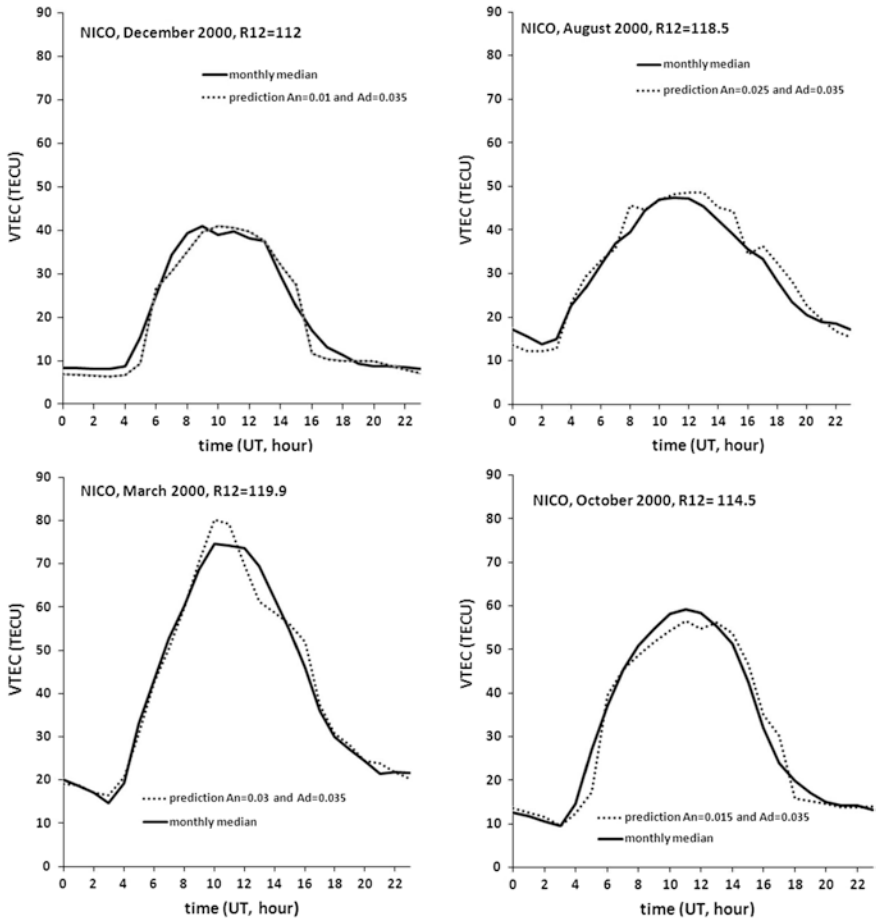


**Fig. 7.3** Comparisons of the observed *VTEC* monthly median at NICO (35.1 N, 33.4 E) in 2000 for December (*top left*), August (*top right*), March (*bottom right*), and October (*bottom left*) with the corresponding predictions

latest version of the NeQuick ionosphere electron density model, developed at the Aeronomy and Radiopropagation Laboratory (now T/ICT4D Laboratory) of the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste, Italy with the collaboration of the Institute for Geophysics, Astrophysics and Meteorology of the University of Graz, Austria and described by Radicella (2009). It is important to emphasize that the Galileo Single Frequency Ionospheric Correction Algorithm is based on the NeQuick ionospheric model.

As described in Chap. 5, the NeQuick model gives the electron density for positions in the ionosphere with height, geocentric latitude, geocentric longitude as coordinates on a spherical Earth. The model values depend on solar activity (given by monthly mean sunspot number $R_{12}$ or 10.7 cm solar radio flux *F10.7*),

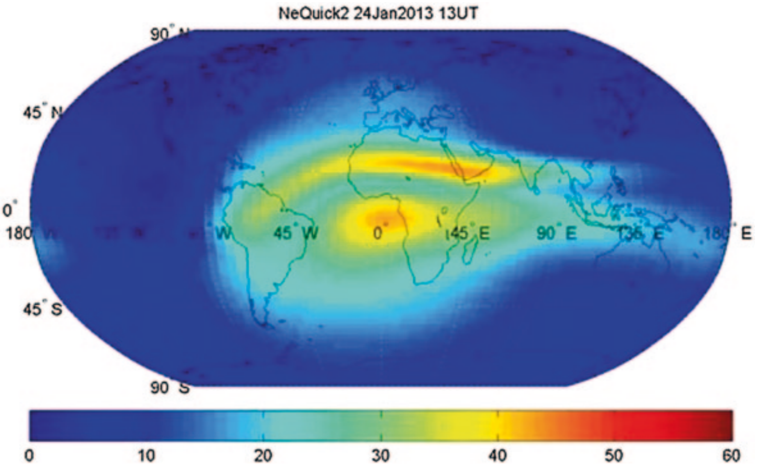


**Fig. 7.4** An example of the *TEC* map computed with NeQuick (Courtesy S. Radicella)

season (month), and time (Universal Time UT). The NeQuick package includes routines to evaluate the electron density along any ground-to-satellite straight line ray path and the corresponding Total Electron Content (*TEC*) by numerical integration (see Fig. 7.4).

In general, they all reproduce ionospheric climatology very well, but as yet cannot reproduce the major ionospheric weather features, shown in Fig. 7.5 as a simple example.

Ground-based geomagnetic field measurements show that two sudden storm commencements occurred on 10 and 11 August 2000 at 0500 UT and 1845 UT, respectively. The maximum *Kp* value reached magnitudes of 7 and 8 and lasted for a few hours during 12 August when *Ap* was 123, *Ri* was 170, and *Dst* was −235 nT (http://www.ngdc.noaa.gov/stp/geomag/geoib.html). These data indicate that this was a major global geomagnetic storm event with, according to Fig. 7.5, a very localized impact on associated *VTEC* variations over the period 10–15 August 2000. Daily *VTEC* values of very different magnitudes understandably do not follow the predicted monthly medians at both HERS and NICO but more importantly there is a significant difference in the nature of their *VTEC* responses to the disturbed solar-terrestrial conditions. At HERS, the ionospheric storm produced a negative phase (daily values well below the monthly median) followed by a clearly defined positive phase (daily values above the monthly median) and then a slow recovery phase, while at NICO it produced a brief initial positive phase followed by negative phase.

Ionospheric storms, as one of the most important phenomena of space weather, and associated forecasting models and methods are described in Chap. 8. The example shown in Fig. 7.4 is used as a simple illustration of the complexity of the problem and the need for total electron content mapping on global, regional, and local scales. It is obvious that the spatial evolution of ionospheric storms is best

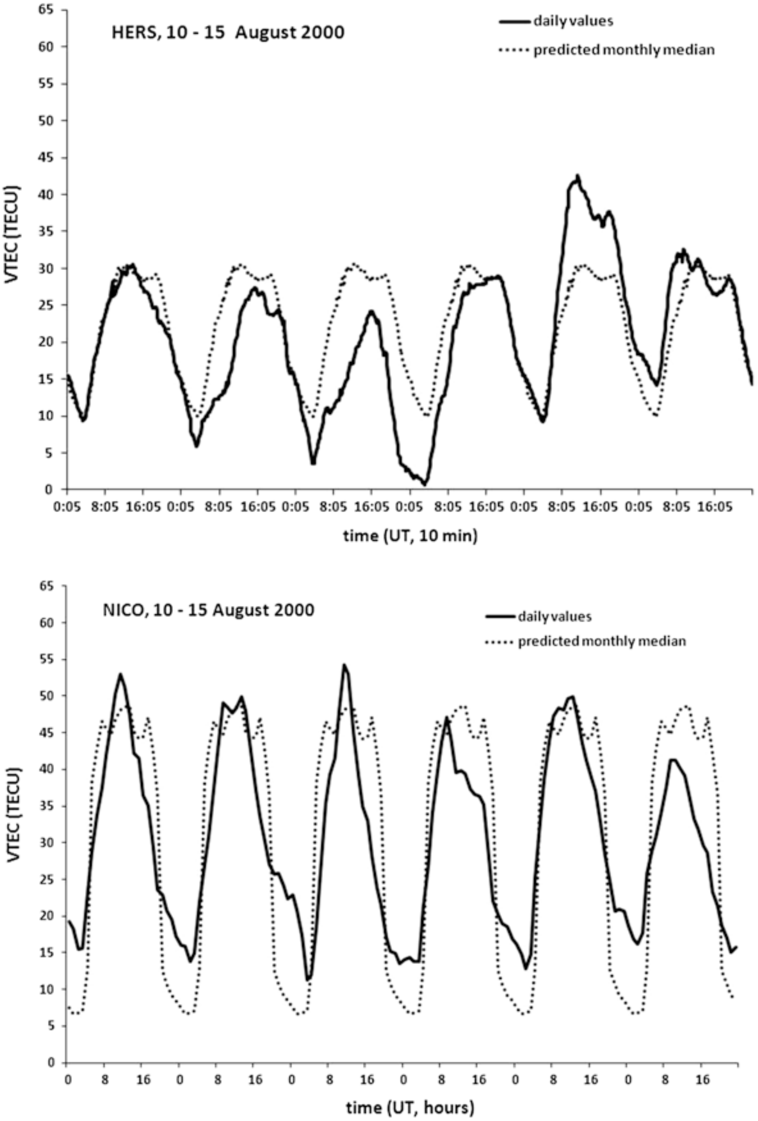


**Fig. 7.5** Daily and predicted monthly median *VTEC* values at HERS (50.9 N, 0.3 E) (*upper panel*) and at NICO (35.1 N, 33.4 E) (*lower panel*) during 10–15 August 2000 when an ionospheric storm was in progress

observed from real-time *TEC* maps generated from GPS reference station data, as provided by services like the IGS. However, these maps require as much *TEC* data as possible in a homogeneous grid in combination with sophisticated and highly accurate mapping techniques.

## 7.4 Global Total Electron Content Mapping

The International GNSS Service (IGS), formerly the International GPS Service, is a voluntary federation of more than 200 agencies worldwide that pool resources and permanent GPS & GLONASS stations (Fig. 7.6) to provide the highest quality data and products, like the standard for Global Navigation Satellite Systems (GNSS) in support of Earth science research, multidisciplinary applications, and education. Consequently the IGS *VTEC* maps have been generated without interruption since 1998 and have been widely disseminated throughout the research community.

The IGS *VTEC* maps are generated in the following steps:

1. Raw GNSS data measurements: provided by the IGS ground network (http://www.igs.org);
2. Independent computation of *VTEC* maps by analysis centres: such as (1) CODE (Centre for Orbit Determination in Europe (http://www.aiub.unibe.ch/content/research/gnss/code_research/index_eng.html), (2) JPL (JetPropulsion Laboratory) USA (http://iono.jpl.nasa.gov/), and (3) gAGE/UPC in Catalonia (http://www.gage.es);
3. Evaluation of the different global *VTEC* maps is provided by the analysis centres. Currently UPC is providing the evaluation, based on their ability to reproduce observed *STEC* variations;
4. A validation with independent sources of *VTEC* data is performed by the Ionospheric Associated Validation Centres (IAVCs) (http://www.uwm.edu.pl/en/);
5. Once the IGS *VTEC* maps are produced they are published on the main IGS distribution server at ftp://cddis.gsfc.nasa.gov/gps/products/ionex/YEAR/DOY.

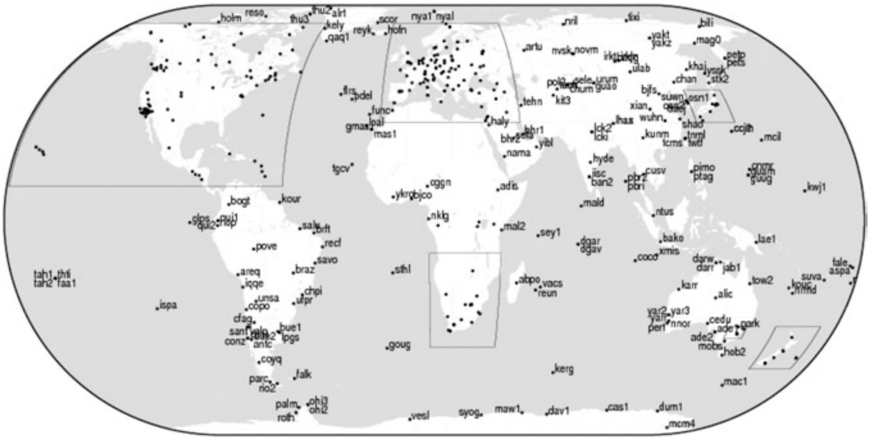


**Fig. 7.6** The IGS tracking network (http://igscb.jpl.nasa.gov/network/complete.html). The close-up maps in boxed regions are generated by dense networks of GPS stations

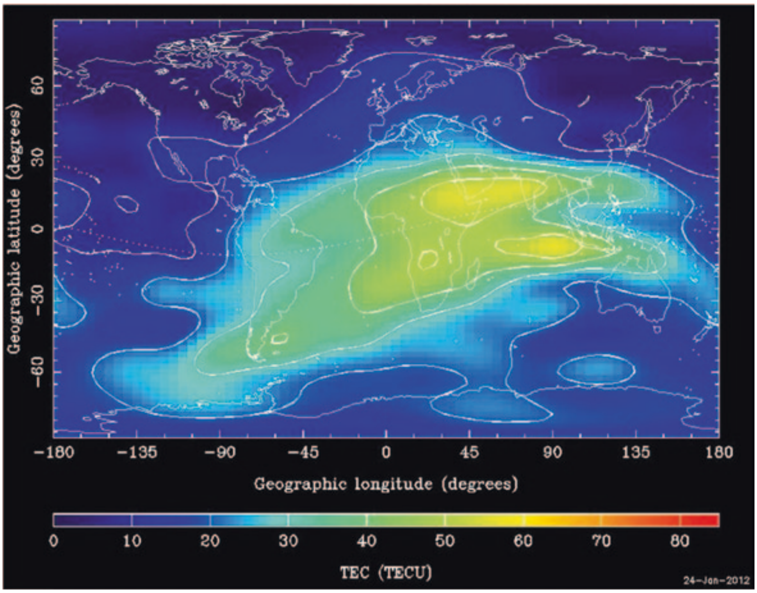


**Fig. 7.7** CODE's global ionospheric map for 19 January 2012 at 1200 UT

A number of global and regional mapping methods currently in use are based on the processing and analysis centre's specific requirements and objectives. An example of CODE's global ionospheric map for 19 January 2012 at 1200 UT is shown in Fig. 7.7.

## 7.5 Regional Total Electron Content Mapping

It is obvious that the worldwide network of GPS receivers provides a unique opportunity to continuously monitor the ionosphere on a global scale. For monitoring ionospheric behaviour on a near real-time basis, a regional approach is more realistic since the majority of receiver locations are within specific areas (see Fig. 7.6).

For the European regional mid-latitude ionosphere, the maps generated by the Kriging spatial-interpolation mapping method are able to estimate the diurnal curve of vertical *TEC* over a location to within 2–3 TECU. Although the number of data points affects the estimation, it should be noted in Fig. 7.8 that more points alone do not necessarily result in a better estimation. Therefore, it is essential to always bear in mind that the ionospheric conditions at certain times and places dictate the map structure (also see Chap. 6).

The advantage of limited area total electron content observations using the local GPS receiver network from the Ordnance Survey UK (OS) lies in its high spatial and temporal resolution (Fig. 7.9). An example of an area-centred ionospheric variability study as seen from the OS GPS receiver network is shown in Fig. 7.10. It further

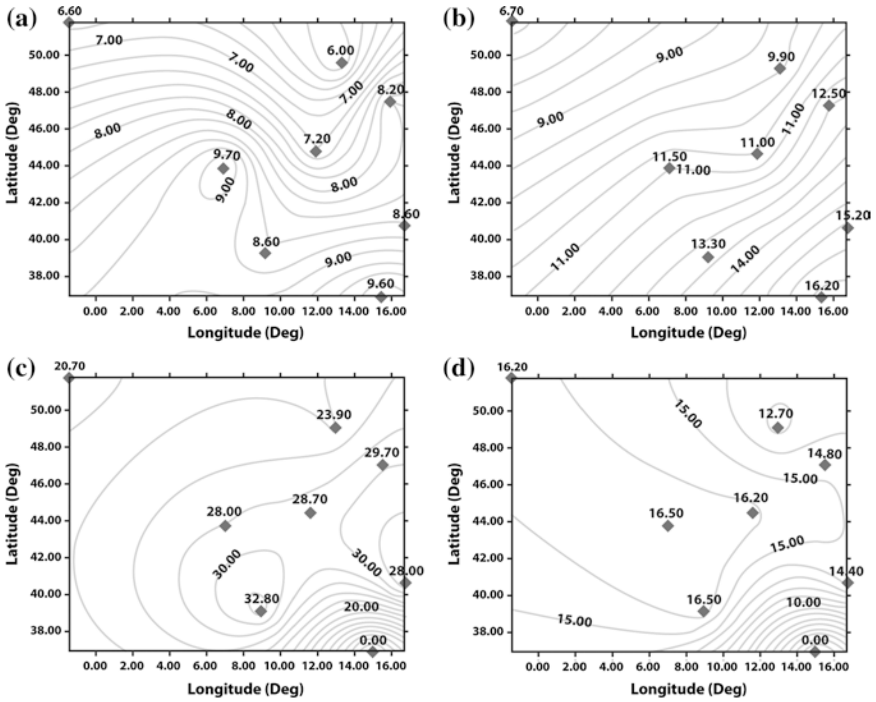
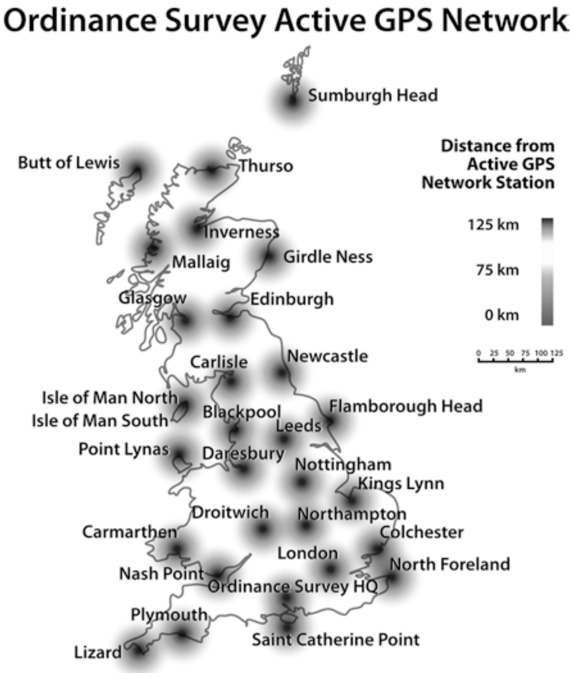


**Fig. 7.8** Regional European ionospheric *VTEC* distribution for 9 October 1998 generated using the Kriging spatial-interpolation mapping method with: **a** data fit for 0000 UT; **b** data fit for 0600 UT; **c** data fit for 1200 UT; **d** data fit for 1800 UT. Contours are labelled in TECU = $10^{16}$ el/m$^2$ and the diamonds indicate measured *VTEC* values at eight GPS receiver locations used for map generation

underlines the relationship between regional plasmaspheric-ionospheric structures and associated solar-terrestrial events. Such studies should lead to a regional storm forecasting technique based on continuous rapid measurements over limited areas similar to the OS. It is also very important for determining the required geographical spacing of *VTEC* monitoring locations for different practical applications.

Kriging spatial-interpolation contouring of the GPS *VTEC* data from all stations shown in Fig. 7.9 was performed for selected days in March and April 2002. The results are presented only as snapshots of daily maps for 31 March and 5 April 2002 at 0005 UT and 1205 UT over the OS area, in Fig. 7.10. For solar-terrestrial conditions in March and April 2002 see SIDC (RWC-Belgium) at http://sidc.oma.be. The maps differ significantly by several TECU, which is expected from the structures in *VTEC* variations, with 31 March under disturbed ionospheric conditions, and 5 April under quiet conditions, affecting the plasmasphere-ionosphere over the area in such a way that the *VTEC* variation appeared as a negative-value phase. More importantly the common pattern consists of

**Fig. 7.9** Map of regional GPS receiver networks from the ordnance survey (OS), UK



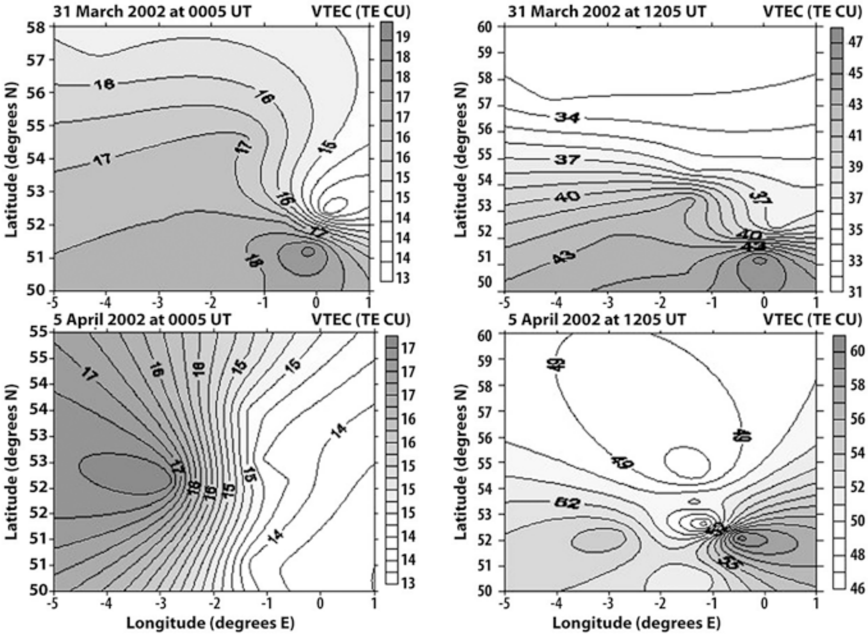


**Fig. 7.10** Maps showing *VTEC* over the OS area for 31 March and 5 April 2002 generated using the Kriging spatial-interpolation mapping method with data fit for 0005 UT (*left panel*), and data fit for 1205 UT (*right panel*)

significant *VTEC* variability over limited geographical areas, indicating the paramount importance of a minimum worldwide grid network for realistic specification of the ionosphere, over and above any given solar-terrestrial conditions.

## 7.6 Total Electron Content and Ionospheric Tomography

An important application of total electron content measurements, developed in recent years as an aid to practical radio systems, is radio tomographic imaging of ionospheric electron density. This involves a recent extension of radio tomographic inversion, using data from a network of GPS receiver stations to create moving images of the ionosphere in three spatial dimensions through time. The radio tomographic method involves measurement of slant electron content, the line integral of electron density, along ray paths from a satellite to a chain of ground stations approximately aligned with the satellite orbital track. Inversion of the data set in a reconstruction algorithm yields a pixellated image in two dimensions of the spatial distribution of the electron density throughout the region of the ray path intersections.

The MIDAS (Multi-Instrument Analysis Technique) inversion method was developed by Mitchell and Spencer (2003) allowing for inversion of the greatly increased quantity and angular coverage from GPS signal measurements, monitored in a widely distributed network of receiving locations. The technique uses measurements of slant electron content obtained from the dual-frequency differential phase observations of GPS satellites in a network of receivers, to produce hour-long animations of electron concentration. This essentially 4-D inversion based on a linear least-squared approach, uses an algorithm incorporating *a priori* information about the evolution of the electron concentration during a specified period of time.

An example of output from the MIDAS method is shown in Fig. 7.11. The left panel shows the *VTEC* over Europe estimated by vertical integration of the

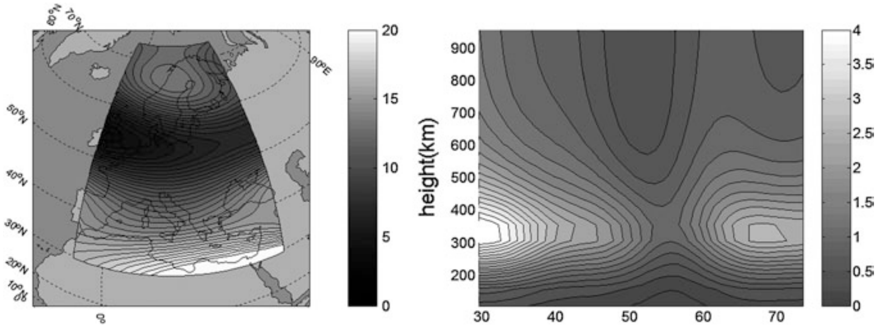


**Fig. 7.11** Frame from an animation of vertical *TEC* (in TECU) over Europe obtained from the inversion of GPS data recorded at about 22 UT on 12 August 2000 (*left panel*) and the corresponding plot of electron concentration ($\times 10^{11} \mathrm{m}^{-3}$) at 15°E longitude (*right panel*) (from Kersley et al. 2004)

electron concentration in the image of one particular frame. A longitudinally extended depletion, corresponding to the main trough, can be seen over Northern Europe at this time. The right panel shows the corresponding height *versus* latitude plot of the electron concentration along longitude 15°E. It can be seen that the location of the trough minimum was well to the south, at about 55°N, consistent with the significant geophysical disturbance on 12 August 2000 (also see Fig. 7.5).

## Suggested Readings

Ciraolo L (1993) Evaluation of GPS L2-L1 biases and related daily TEC profiles. In: Proceedings of the GPS/Ionosphere Workshop, Neustrelitz, pp 90–97

Davies K, Smith EK (2000) Ionospheric effects on land mobile satellite systems. In: Goldhirsh J, Vogel WJ (eds) A supplement to "Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results". NASA, Reference Publication 12, University of Colorado, Boulder

Dow JM, Neilan RE, Rizos C (2009) The international GNSS service in a changing landscape of global navigation satellite systems. J Geod 83:191–198

Hernández-Pajares M, Juan JM, Sanz J, Orus R, Garcia-Rigo R, Feltens J, Komjathy A, Schaer SC, Krankowski A (2009) The IGS VTEC maps: a reliable source of ionospheric information since 1998. J Geod 83:263–275

Hernández-Pajares M, Miguel Juan J, Sanz J, Aragón-Àngel À, García-Rigo A, Salazar D, Escudero M (2011) The ionosphere: effects, GPS modeling and the benefits for space geodetic techniques. J Geod 85:887–907

Jakowski N (1996) TEC monitoring by using satellite positioning systems. In: Kohl H, Ruster R, Schlegel L (eds) Modern Ionospheric Science. European Geophysical Society, Katlenburg-Lindau

Jakowski N, Hoque MM, Mayer C (2011) A new global TEC model for estimating transionospheric radio wave propagation errors. J Geod 85:965–974

Kersley L, Malan D, Pryse SE, Cander LjR, Bamford R, Belehaki A, Leitinger R, Radicella SM, Mitchel CN, Spencer PSJ (2004) Total electron content—a key parameter in propagation: measurement and use in ionospheric imaging. Ann Geophys 47:1067–1091

Mitchell CN, Spencer PSJ (2003) A three dimensional time-dependent algorithm for ionospheric imaging using GPS. Ann Geophys 46:687–696

Radicella SM (2009) The NeQuick model genesis, uses and evolution. Ann Geophys 52:417–422

Ventura-Traveset J, Flament D (eds) (2006) EGNOS: the European Geostationary Navigation Overlay System-A Cornerstone of Galileo. Eur. Space Agency Spec. Publ., ESA SP-1303

Vernon A, Cander LjR (2002) Regional GPS receiver networks for monitoring local mid-latitude total electron content. Ann Geophys 45:649–656

# Chapter 8
# Ionospheric Forecasting

## 8.1 Introduction

A new era in the history of ionospheric forecasting began during the International Geophysical Year (IGY, July 1957 to December 1958), based on earlier efforts to predict ionospheric effects on HF radio communications. Solar-terrestrial physics had developed from the traditional disciplines of ionospheric physics, geomagnetism, solar physics, and cosmic ray physics, and was to assume a prominent role with the inauguration of World Data Centers (WDCs). WDCs for Solar-Terrestrial Physics (STP) in Boulder, Moscow, Chilton, and many other locations began systematically collecting, analyzing, filing, and distributing worldwide the routinely observed geophysical data through space and time, describing the environment between the surface of the Sun and the surface of the Earth.

These data are available on various supports including CD-ROMs, Internet with real-time on-line access via the World Wide Web, and other supports on request. Associated nationally and internationally coordinated research efforts over the last few decades have greatly increased scientific understanding of the dynamic plasma processes that create a disturbed space environment, with significant impact on our contemporary technological society (Sect. 8.2). This mainly applies to ionospheric forecasting for modern terrestrial and Earth-to-satellite telecommunications, navigation, and positioning systems during disturbances to the Earth's environment (Sect. 8.3).

In addition to the diurnal, seasonal, and solar cycle variations described in the previous Chapters, the ionosphere experiences obvious day-to-day variations under both geomagnetically quiet and disturbed conditions. Considering that the observed daytime standard deviations are typically less than ±6 % of the monthly median value for *foE* and *foF1* and about ±15 % for *foF2* at mid-latitudes, and that during ionospheric disturbances average deviations sometimes go beyond ±25 % for *foF2*, ±10 % for *foF1*, and ±4 % for *foE*, most attention is paid to F region disturbance during geomagnetic storms. In particular its *foF2* and *M(3000)F2* characteristics, these being the most important for radio communications, and *TEC* for

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_8,

navigation and positioning. Other perturbations, particularly in the D and E regions, are briefly mentioned here to complete the picture of this phenomenon (Sect. 8.2).

A forecast is a prediction for a specific time in order to anticipate the day-to-day variability in the Earth's ionosphere. To develop such a prediction, advances in ionospheric research must be integrated with applied research in forecasting. However, the present imperfect knowledge of the effects of solar-terrestrial processes, even on the mid-latitude ionosphere, limits accurate forecasting significantly for many practical applications. The traditional ionospheric forecasting methods developed up to now depend on successful solar and magnetospheric forecasts (Sect. 8.3.1). The nonlinearity of this prediction issue led to the use of the time series prediction capabilities of artificial neural networks, among other artificial intelligence techniques, for a solution (Sect. 8.3.2). Short-term ionospheric prediction over particular regions is a novel operational forecasting tool available in recent years on the World Wide Web for interactive use based on continuous monitoring of ionosphere conditions in near or real-time. Associated forecasting mapping techniques are therefore described in Sect. 8.3.3.

## 8.2 Ionospheric Disturbances

The effects of solar phenomena and associated geomagnetic, magnetospheric, and ionospheric activity on near–Earth plasmas vary in importance, timescale, predictability, and consequences (Table 8.1). They can be (1) direct effects, caused by rapid changes in solar UV radiation and X-ray illumination of the Earth's ionosphere and atmosphere during solar flares, and (2) indirect effects, caused by complex interactions between the solar wind and the coupled magnetosphere-ionosphere-atmosphere system. In the latter case, the level of disturbance of the geomagnetic field serves as a convenient proxy for geomagnetic storms, indicating variations in the intensity of the Earth's magnetic field, which may reach several percent of the undisturbed value measured on the surface.

### *8.2.1 Sudden Ionospheric Disturbance (SID)*

The intense associated short-wave electromagnetic radiation during a major solar flare produces strong ionization of the D region. Increased D region electron density by day with higher density of neutral particles results in enhanced HF radio

**Table 8.1** Types of solar emission

| Solar emission | Time of transit to Earth | Terrestrial consequences |
|---|---|---|
| X-rays | 8.3 min | Sudden ionospheric disturbance |
| Protons, 1–100 MeV | A few hours | Solar proton event |
| Low energy plasma, velocity ~ 1000 km/s | 1–2 days | Geomagnetic storm and aurora |

wave absorption. The phenomenon was originally discovered in short-wave radio signal attenuation in radio communications and is also known as the Mogel-Dellinger effect. Similar attenuation occurs for cosmic noise of extraterrestrial origin passing through the D region.

This Sudden Ionospheric Disturbance (SID) or Shortwave Fadeout (SWF) effect immediately follows a solar flare and appears globally on the sunlit side of the Earth. SID events have an onset of a few minutes and can last up to an hour. The electron density increase in the D region may be as high as an order of magnitude (or even higher) in response to a flare of X-rays $\lambda < 10$ Å. An increase in electron density by about 20–30 % in response to solar flares is also observed in the E region for X-rays $\lambda > 10$ Å. Solar flare effects are minor at F region heights, due to ionization from EUV radiation.

Since the rate of solar flare occurrence increases and decreases in coordination with the sunspot cycle, navigators find it more difficult to use Loran-C and Omega systems during daylight hours during the maximum phase of the solar cycle, while GNSS signals are generally immune to ionospheric changes in response to large infusions of solar X-rays.

### *8.2.2 Polar Cap Absorption*

The arrival of high-energy protons with $E \geq 5$–30 MeV released from the Sun at the time of a flare can also produce additional ionization in the ionospheric D region of the polar caps. It persists from several hours to a few days and often results in a breakdown of MF/HF communications on polar routes. This activity is more pronounced during years near to solar maximum and affects the polar latitudes only, and so solar proton events (SPE) are often called Polar Cap Absorption (PCA) events. They occur during both daytime and night-time and are possibly the most important disturbance of the high-latitude lower ionosphere, affecting both ground-based and space-based systems. SPE also present a potential hazard for space travel and monitoring and forecasting are considered an important part of space weather.

### *8.2.3 Ionospheric Storm*

The ionospheric response to geomagnetic storms, one of the most dramatic consequences of magnetosphere-ionosphere-thermosphere coupling, with changes in solar wind conditions, is of long-standing and very contemporary interest. These phenomena are known as ionospheric storms and tend to generate large disturbances in density distribution, total electron content, and the current system. Extensive studies over the years have highlighted the dynamic character of ionospheric storms and established evidence for better understanding the physical mechanisms that control their morphology and phenomenology.

F region ionospheric storms result from a input of solar wind energy captured by the Earth's magnetosphere and then released and dissipated into the auroral ionosphere, which establishes a complex morphology of temperature, winds, electric fields, and composition changes. These continue for a few hours to several days and lead to significant changes in ionospheric plasma parameters, during which time the band of available frequencies for radio waves propagated by ionospheric reflection is reduced.

Although ionosonde observations are subject to limitations, the characteristic behaviour of ionospheric storms was first established by means of ionosonde measurements at mid-latitudes, measuring *foF2* and $h'F2$, and by trans-ionospheric radio propagation, measuring *TEC* as the most significant parameters. In general, large perturbations of ionization density at F region height during storms are either positive or negative, with *foF2* significantly reduced relative to a quiet reference level for negative storms, and increased for positive storms.

Numerous investigations have been made of how to quantify deviation from quiet ionospheric reference values as a function of geomagnetic activity, month, and solar epoch, for use in operational forecasting. One of the many problems, is that there is still no international agreement among ionospheric scientists of a quiet reference level. The most common proposal, described in this book, is the use of monthly median values. This seems sensible and does not exclude different approaches such as monthly mean, running mean, weighted means of preceding days, geomagnetically quietest days (Q-days defined in Sect. 2.2.3), etc. Another problem is related to the various attempts to find any reliable statistical correlation between daily ionospheric characteristics, such as *foF2* and *M(3000)F2* values, with the conventional indices of solar and/or geomagnetic activity described in Sect. 2.2.3.

In order to demonstrate ionospheric storm effects on peak electron density *NmF2* and vertical total electron content *VTEC* at mid-latitudes, a specific case is considered, the storm of 5 April 2010. Ground geomagnetic field measurements show that a sudden storm commencement (SSC) occurred on 5 April 2010 (D1 day) at 0826 UT with the maximum *Kp* value reaching 8 and $Ap = 55$. According to the Royal Observatory of Belgium at http://www.sidc.be/sunspot-data the source of the disturbance was a halo CME (Coronal Mass Ejection), which occurred on 3 April 2010. A strong interplanetary shock was observed by the NASA's ACE spacecraft at 0756 UT on 5 April. Conditions returned to normal after a few days, with 10 April as Q2 day in Fig. 8.1 representing a reference quiet level.

Figure 8.1 shows a good example of positive and negative storm effects on ionospheric peak electron density *NmF2* and vertical total electron content *VTEC* at the Chilton ionosonde station and co-located GPS station HERS respectively, illustrating the following simplified summary of the major phenomena of F region storms at mid-latitudes: (1) After a sudden storm commencement (SSC), the peak electron density *NmF2* and total electron content *VTEC* increase for a few hours; this is the initial positive phase, (2) Several hours later the main phase sets in taking the form of a severe depletion of *NmF2* and *VTEC*; this is a negative storm. If *NmF2* decreases to the extent that $NmF2 < NmF1$ this is the so-called G-condition;

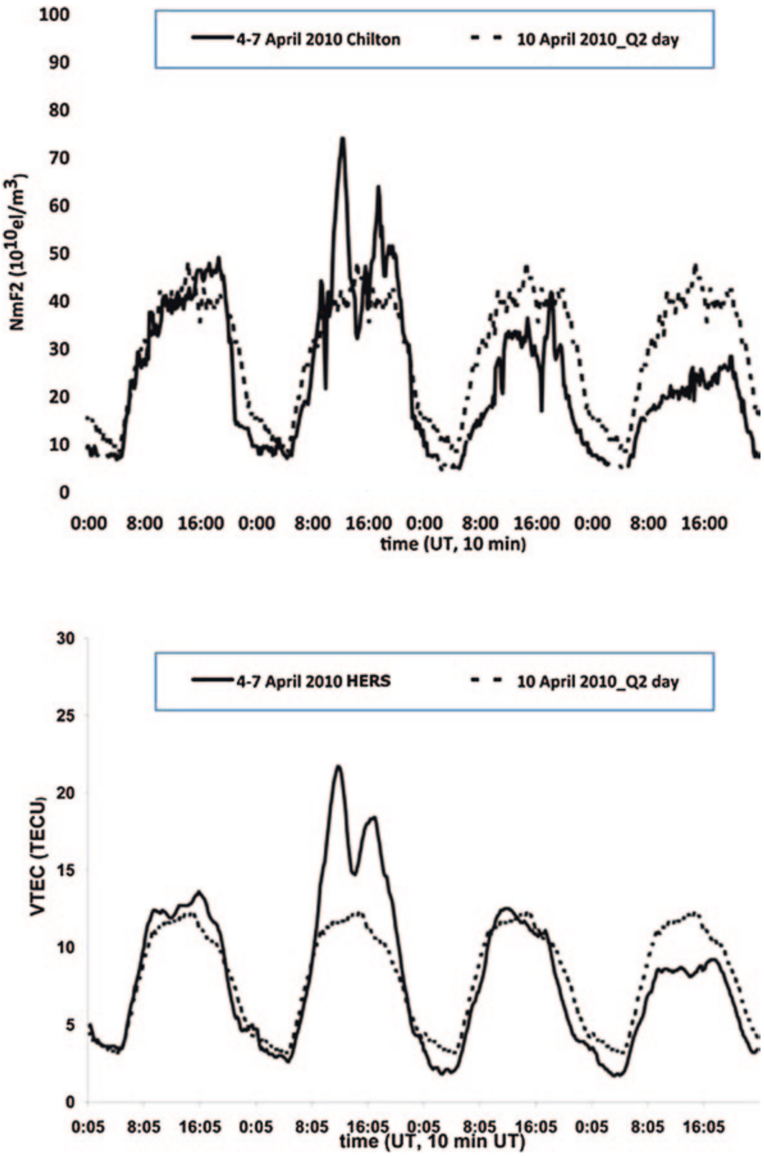


**Fig. 8.1** Storm effects on ionospheric peak electron density *NmF2* at the Chilton (51.6° N, 358.7° E) ionosonde station (*upper panel*) and vertical total electron content *VTEC* at the co-located GPS station HERS (50.9° N, 0.3° E) (*lower panel*)

(3) Sometimes, especially in the ionospheric winter season, *NmF2* and *VTEC* are enhanced in the main phase; this is a positive storm, and (4) Using the ratio of electron content to maximum electron density, which defined slab thickness with Eq. 3.5.3, it can be demonstrated that during storms the shape of the *N(h)*

distribution below the F2 peak changes in such a way that the real height of the layer increases. However, this is by far less than the virtual height $h'F2$ increase frequently observed on ionograms.

The origin of negative ionospheric storms is well documented and attributed to changes in the neutral gas composition of the upper atmosphere. On the other hand, it is widely reported in a number of studies that thermospheric winds and electromagnetic fields are the two main drivers of positive ionospheric response to geomagnetic storms. Positive ionospheric storms are caused by upward transport of ionization density, which is the immediate cause of the increased density. However, the question as to which of these two plays a dominant role in ionospheric electron density changes from high-latitude regions (>60°) through mid- and low-latitudes to equatorial regions is still under discussion.

In addition, the great variability in the environment paradoxically shows that from time to time a quiet geomagnetic field does not necessarily mean an undisturbed ionosphere. It has been shown that storm-like variations, scattered enhancements or depletions during undisturbed geomagnetic conditions, are not only confined to low-latitudes and equatorial regions. As a result, some storm effects at all latitudes and longitudes are not easy to distinguish from large quite-time variations and even more difficult to forecast using traditional or any other methods.

In conclusion, it should be noted that, in addition to forecasting, ionospheric storm studies are important from at least three separate standpoints. Firstly, they provide an improved understanding of the physical processes involved in energy deposition and ionospheric movements. Secondly, they enable extreme ionospheric conditions to be quantified so that likely variability bounds can be defined, of particular importance when planning telecommunications. Thirdly, they assist investigations into the spatial correlations of ionospheric change under different conditions, which is important for the development of instantaneous mapping algorithms as mentioned in Chap. 6.

## 8.3 Ionospheric Forecasting Techniques

Numerous case studies, simulations and theoretical results provide evidence that any algorithm for short-term ionospheric forecasting should include at least three components: (1) solar and geomagnetic activity forecasting, (2) ionospheric storm forecasting, and (3) modelling of ionospheric characteristics and parameters for storm conditions. In general, short-term ionospheric modelling is usually defined as a quantitative indication of the critical frequency $foF2$ variations up to 24 h ahead.

Specifically, short-term models for radio service frequency management are directed towards assessing the best frequency to use with an existing system, in the light of the prevailing ionospheric conditions of the time. Users generally require near real-time ionospheric forecasting for radar and surveillance applications, and from 1 to 24 h ahead for point-to-point and mobile communications. Therefore, there is special interest in advance storm warnings for major storms for the former, and there are a number of methods available for forecasting characteristics relevant to ionospheric telecommunication applications for the latter.

### *8.3.1 Statistical Methods*

Many techniques have been developed for ionospheric forecasting, among which a widely used empirical approach to short-term F2 layer prediction based on statistical methods including various multiple linear regression approaches, the autocorrelation method, an empirical storm-time ionospheric correction model, etc. For detailed descriptions of a number of specific statistical models, see A.V. Mikhailov et al. in 2007.

Within European ionospheric COST actions, the majority of the forecasting techniques attempt to derive a statistical model for the *foF2* and *M(3000)F2* deviations around the long-term model monthly median values based on hourly *foF2* and *M(3000)F2* data collection for four ionospheric seasons, 24 h of local time, and three levels of solar activity (for instance $R_{12} < 60$, $60 < R_{12} < 120$, $R_{12} > 120$). Using these data it is possible to establish the standard deviation, lower and upper deciles of the *foF2* and *M(3000)F2* deviations for the solar-terrestrial conditions in question. However, an uncertainty of the order of 0.5 MHz exists at all times and, when extrapolating to remote locations where measured data are not available, errors are likely to be prohibitive.

Below, a linear technique is described, based on an autocorrelation method (ACM) developed within the European ionospheric community to predict *foF2* up to 24 h ahead for a number of ionospheric stations, with encouraging results. The ACM is based on a superposition of diurnal periodic components with an exponential term representing short-term decorrelation of the time-series. Although in ionospheric studies, the average diurnal variations of a characteristic are traditionally represented by the monthly medians, in the ACM approach, instead of the ionospheric characteristic itself its relative deviation from the median values is used. In the case of *foF2*, this quantity is given by:

$$\Phi = (foF2 - foF2_{med})\,/foF2_{med} \tag{8.3.1}$$

where $\Phi$ is considered as a random process over time. The advantage of this quantity is that it does not contain any periodic components and can be regarded as a stationary process over an observation period of a few days. Assuming an autoregressive model of the form:

$$\Phi(t_0) = \sum_{k=1}^{n} \beta_k \Phi(t_0 - \tau_k) \tag{8.3.2}$$

where $n$ is the order of the regressive model, and $\beta_k$ for $k = 1, \ldots n$, are weighting coefficients to be determined.

When the process is stationary, if there is a sample of $N$ measured values $\Phi_i$ at times $t_i$ ($i = 1,\ldots, N$) ordered in decreasing time, the weighting coefficients $\beta_k$ are solutions of the system of equations:

$$\sum_{k=1}^{n} \beta_k r_{\Phi\Phi}\left(\tau_{ik}\right) = r_{\Phi\Phi}\left(\tau_i\right) \quad i = 1,2,\ldots,n \tag{8.3.3}$$

where $\tau_{ik} = \tau_i - \tau_k$ is the time shift and $r_{\Phi\Phi}$ denotes the autocorrelation function of $\Phi$. In practice the true $r_{\Phi\Phi}$ is not known, but only the normalized empirical autocorrelation function $\rho_{\Phi\Phi}$, defined by Korn and Korn (1968):

$$\rho_{\Phi\Phi}\left(\tau\right) = \frac{\sum \Phi_i \Phi_{k(i)}}{\sqrt{\left[\sum \Phi_i^2\right]\left[\sum \Phi_{k(i)}^2\right]}} \tag{8.3.4}$$

where the summation in the numerator is taken over the pairs of $\Phi$ values having the same time difference $\tau$ in the data sample.

In addition, an assumption that the empirical mean $\bar{\Phi} = \frac{1}{N}\sum_i \Phi_i$ is zero can be justified for ionospheric series over a period of several days. The solution for Eq. 8.3.3 is adopted as a correlation function of the following form:

$$r_{\Phi\Phi}(\tau) = \exp\left(-\frac{|\tau|}{T}\right) \tag{8.3.5}$$

where $T$ is the time constant of the process $\Phi(t)$. Numerical tests give optimal values for $n$ and $N$ of 24 (1 day) and 600 (25 days) respectively.

To apply this model to 24-hour prediction of *foF2*, the empirical autocorrelation function $\rho_{\Phi\Phi}$ over the 25-day period for values of time lags $\tau_k$ up to 4 days is calculated with estimated typical values for the time constant $T$ of the order of 10 h by least squares filling for the system as:

$$\rho_{\Phi\Phi}(\tau_k) = \exp\left(-\frac{|\tau_k|}{T}\right) \quad k = 1,\ldots,24 \tag{8.3.6}$$

Within the 4-day range of $\tau$, from Eq. 8.3.5 the 24 time lags are selected with the highest autocorrelation values, which are used to obtain the weighting coefficients $\beta_k$ from the system of Eq. 8.3.3. Finally, $\Phi_0$ is obtained from Eq. 8.3.2 and the procedure is repeated for each hour $i$ ($i = 1\ldots24$). An example of one-day *foF2* prediction is shown in Fig. 8.2.

The one-day prediction is performed for each of the days between 21 and 31 July 1981 for *foF2* data from the Slough ionosonde station. The dots represent the measured *foF2* (in tenths of MHz) and the solid line shows the predictions. On 25 July an intense geomagnetic storm started at 1230 UT and during the following night *foF2* dropped to 50 % of its median level. It can be seen that the prediction

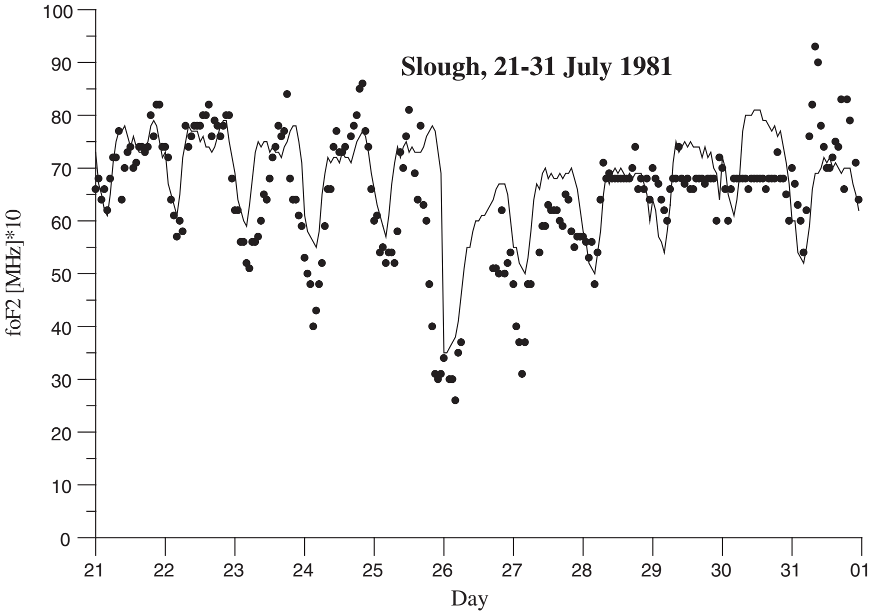


**Fig. 8.2** The one-day prediction of *foF2* (*solid line*) compared with measured data from Slough (51.5° N, 359.4° E) ionosonde station (*dots*) for the period 21–31 July 1981, when an intense geomagnetic storm was in progress from 1230 UT on 25 July 1981

follows this drop fairly well, but with a few hours delay. This reaction time is natural, since values for the few previous hours usually have the highest weighting and therefore make the biggest contribution in Eq. 8.3.2. The ionosphere recovered during the next few days as also predicted by the method.

The Geomagnetically Correlated Autoregression Model (GCAM) is an improved version of the ACM forecasting capabilities obtained by incorporating cross-correlation between the *foF2* and the *Ap* index into the autocorrelation analysis previously considered. GCAM is still an extrapolation model based on the improved model's prediction and driven by current and most recent *foF2* values, estimates of the reference ionosphere, and prediction of a synthetic geomagnetic activity index *G* in hourly resolution based on the daily *Ap* index, for which a 45-day prediction is available at the Space Weather Prediction Center of NOAA, Boulder (http://www.swpc.noaa.gov/). A new term in the regression Eq. 8.3.2 was added to express a linear dependence of Φ on geomagnetic activity, and then the weighting coefficients in the regression equation were obtained through the autocorrelation functions of Φ and *G* and their cross-correlation function.

The model performs consistently well during disturbed ionospheric conditions and therefore the European Digital Upper Atmosphere Server (DIAS) forecasting products and services are currently supported by the implementation of the GCAM as described by I. Tsagouri in 2011.

### *8.3.2 Neural Network Methods*

Dynamic ionospheric forecasting for resource management of telecommunication subsystems, requires accurate real-time knowledge of the top and bottom-side *N*(*h*) profiles. Hence, *foF2* and *TEC* values are required both in real-time mode and in hours-ahead prediction. It was the nonlinearity of this prediction problem that led to application of an artificial neural network (ANN) method as a solution. An artificial neural network is a structure of computing elements which allows easy integration of *a priori* knowledge, as well as having the ability to represent a complex dynamic system through training with data.

Early successful attempts to build various artificial neural network models for ionospheric critical frequency *foF2* prediction and forecasting were based on monthly and daily time scales. Further development included a number of studies related to forecasting GPS *TEC* using the neural network technique, neural network development for the forecasting of upper atmosphere parameter distributions, a neural network-based ionospheric model for the auroral zone, a new neural network-based global F2 peak electron density model for the International Reference Ionosphere (IRI), a comparison between the ANN forecast *foF2* values and analytically calculated IRI-based *foF2* values, and so on.

A simple way to introduce a basic idea about artificial neural network in the ionospheric studies is to use a schematic description as in Fig. 8.3. A neural network is composed of several layers of neurons: an input layer, one or more hidden layers and an output layer. Each layer of neurons receives its input from the previous layer or from the network input. The output of each neuron feeds the next layer or the output of the network. The first layer is always an input layer that distributes the inputs to the hidden layer.

The unknown parameters of any neural network are the weights which can be found through training with different algorithms on the known input–output patterns. Therefore, training a neural network involves a database of examples which are values for the input and output. The neural networks would learn by adjusting the weights connecting neurons in different layers to minimise the error of outputs. These algorithms try to minimise the error between the desired output and the network output by adjusting the weights according to gradient descent. The functionality of different neural networks was described in detail by S. Haykin in 1994.

A concise summary of the NNARX—neural network based autoregressive model with additional inputs (*X*)—is outlined below. This can provide forecasting of the hourly variation of *foF2* and *TEC* for 1-hour ahead, as the shortest time

**Fig. 8.3** The modular neural network

scale over which the ionosphere can be predicted forwards extremely well. The forecast periods were chosen as February 1986 for *foF2* and December 1990 for *TEC*, these representing low and high solar activity respectively, under various geomagnetic conditions.

The NNARX model was trained using hourly *foF2* values from 1 January to 31 December of 1986, excluding February 1986, the month for which forecasting was conducted. The forecast period was chosen as February 1986, this being a month of low solar activity during which a severe geomagnetic storm occurred on 6 February. Table 8.2 and Fig. 8.4 give some details concerning the solar-terrestrial conditions during selected periods.

In Table 8.2 *Ri* is the international sunspot number and *Ap* is an averaged planetary geomagnetic index based on data from a set of specific geomagnetic observatories (see Sect. 2.2.3). It is customary to use the storm classification as follows: (1) Minor Geomagnetic Storm: a storm for which the *Ap* index was greater than 29 and less than 50; (2) Major Geomagnetic Storm: a storm for which the *Ap* index was greater than 49 and less than 100; (3) Severe Geomagnetic Storm: a storm for which the *Ap* index was 100 or more. The *Dst* index is an index of geomagnetic activity derived from a network of near-equatorial geomagnetic observatories that measure the intensity of the globally symmetrical equatorial electrojet (the "ring current"). Peak *Dst* values are between −20 and 20 nT during quiet geomagnetic

**Table 8.2** Solar-terrestrial conditions during storm periods in February 1986 and December 1990

| Months | Day of storm commencement | Min *Dst* index | Max *Ap* index |
|---|---|---|---|
| February 1986 with monthly mean sunspot number $Ri = 23.2$ | 6 February at 1312 UT | −307 nT on 9 February | 208 on 8 February |
| December 1990 with monthly mean sunspot number $Ri = 129.7$ | 21 December at 1724 UT | −47 nT on 24 December | 15 on 24 December |

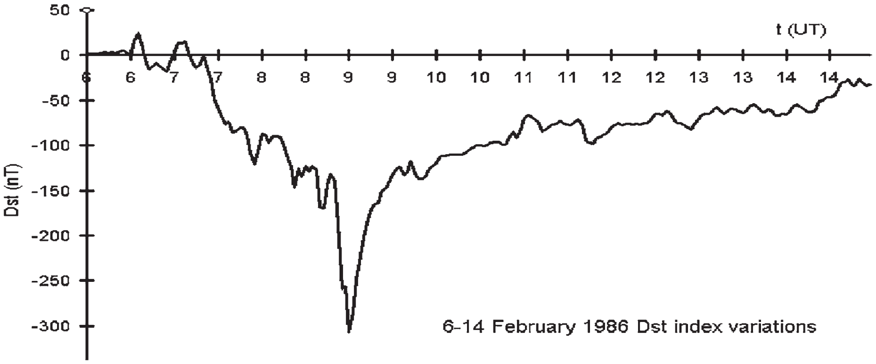


**Fig. 8.4** *Dst* variations during the 6–14 February 1986 storm

conditions and decrease to below $-100$ nT during highly disturbed periods. In the case of severe geomagnetic storms, *Dst* values are below –250 nT.

NNARX—neural network based autoregressive model with additional inputs *(X)*—is one possible approach that uses the hybrid time-delay multi-layer perception neural network with only the critical frequency of the F2 layer as input parameter, to produce one output *foF2* value at hour $t+1$. Inputs *(X)* include the *foF2* value at time *t*, seven days mean *MfoF2* values and appropriate differences *Delta MfoF2* at particular selected hours (*t, t–1, t–23, t–47*) as well as *MfoF2* at forecast time $(t+1)$ calculated using only the learning set of data to generate the background daily variations of *foF2*. The NNARX is trained using a cost function with a weight-decay term. Experimental experience has shown that the weight-decay scheme alone is not enough to overcome the overfitting problem. Therefore a pruning scheme is used to remove superfluous parameters from the network architecture. The so-called Akaike's Final Prediction Error (FPE) estimate of the test error then determines which structure is optimal. The data for the *foF2* forecasting 1-hour ahead are values taken from the Rutherford Appleton Laboratory CD-ROM produced within the COST 251 action (http://www-cost251.rcru.rl.ac.uk/) for a few European ionosonde stations listed in Table 8.3.

The forecast accuracy is established by clearly stating the range and scope of test conditions and showing the breakdown of training and test data, as well as indicating any comparisons made with other or reference models, like for example persistence. For comparative purposes the accuracy of the NNARX method is quoted in terms of the Normalised Root Mean Square Error (*NRMSE)* defined as:

$$NRMSE = \frac{1}{\sigma}\sqrt{\frac{\sum_i (x_i - y_i)^2}{n}} \tag{8.3.7}$$

where $x$ are the actual values, $y$ are the predicted values, $n$ is the total number of comparisons between actual and predicted values and $\sigma$ is the standard deviation of the actual values from the mean value, $X$, given by:

$$\sigma^2 = \frac{\sum_i \left(\overline{X} - x_i\right)^2}{n} \tag{8.3.8}$$

**Table 8.3** *RMS* and *NRMS* errors with neural network and persistence at some of the European ionosonde stations for February 1986

| Ionosonde stations | *RMSE* (MHz) | | *NRMSE* | |
|---|---|---|---|---|
| February 1986 | Persistence | Neural network | Persistence | Neural network |
| Slough SL015 (51.5° N, 359.4° E) | 0.6482 | 0.4456 | 0.4740 | 0.3258 |
| Poitiers PT046 (46.6° N, 0.3° E) | 0.6178 | 0.4460 | 0.4469 | 0.3226 |
| Rome RO041 (41.9° N, 12.5° E) | 0.8097 | 0.4745 | 0.3568 | 0.2091 |
| Sofia SQ143 (42.7° N, 23.4° E) | 0.7213 | 0.5258 | 0.5285 | 0.3853 |
| Moscow MO155 (55.5° N, 37.3° E) | 0.6121 | 0.3778 | 0.4288 | 0.2647 |

The predictive accuracy of the model for all stations is summarised in Table 8.3. All errors are also quoted using the Root Mean Square Error *(RMSE)* in MHz, with respect to the one step ahead NNARX predictive model, for the entire test set in each instance. The persistence model is included as an effective reference in each case.

Encouraging results in Table 8.3 show that there is a predictive advantage to neural network modelling which makes it a useful tool for introduction into current ionospheric space weather applications. For these purposes, the effects of different solar-terrestrial indices on 1-hour ahead ionospheric forecasting using the NNARX model are presented in Table 8.4. Usually the accuracy of a forecasting procedure is given in terms of relative error, when different input parameters have been used, as follows:

$$RE = \frac{1}{N} \sum_{i=1}^{N} \frac{abs(y_i - x_i)}{y_i} \cdot 100\,\% \tag{8.3.9}$$

A more in-depth investigation of neural network forecasting of the key ionospheric characteristic *foF2* at different European ionosonde stations (Table 8.4) clearly demonstrates that one hour ahead forecasting performance is not significantly improved by adding solar activity *Ri* and geomagnetic *Ap* and *Dst* indices. It should be emphasised that data and comparisons are specifically for a few stations but the results are applicable to the entire mid-latitude region of the European longitudinal domain. Indeed, the NNARX mode requires neither predicted nor past values of solar and geomagnetic indices to forecast *foF2* values one hour ahead even during very disturbed geomagnetic periods, like that of February 1986 (see Fig. 8.5).

The NNARX *foF2* neural networks forecasting model can easily be modified to introduce *TEC* prediction one hour ahead. Although extensive comparisons have been made between observed values of *TEC* and those calculated using neural networks, an example is given in Fig. 8.6 for 1-hour ahead *TEC* forecasting during December 1990. This figure illustrates a very small subset of the comparisons that could be made with the available *TEC* data determined from Faraday rotation observations at Florence (43.8° N, 11.2° E) using the signal of the OTS-2 satellite. In this case, a learning *TEC* set covered all hours in 1990 except for 744 h in December 1990, which were used as a test set. Again, the agreement between predictions and measurements range from very good (daytime) to good (night-time).

**Table 8.4** *RE* at all ionosonde stations from Table 8.3 for February 1986

| Ionosonde stations/Input parameters | SL051 *RE* (%) | PT046 *RE* (%) | RO041 *RE* (%) | SQ143 *RE* (%) | MO155 *RE* (%) |
|---|---|---|---|---|---|
| *foF2* only | **8.94** | 9.14 | 8.86 | 10.22 | **8.95** |
| $foF2 + Ri + Ap$ | 9.10 | 8.63 | 8.71 | 10.07 | 9.88 |
| $foF2 + Ri + Dst$ | 9.18 | 8.89 | 8.93 | **9.86** | 9.59 |
| $foF2 + Ri + Ap + Dst$ | 9.26 | **8.60** | **8.70** | 9.90 | 9.97 |

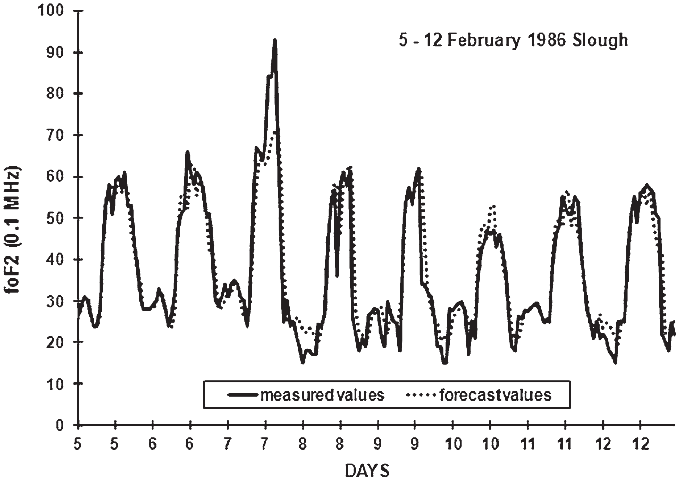


**Fig. 8.5** Measured and forecast *foF2* values for 5–12 February 1986 at Slough (51.5° N, 359.4° E) ionosonde station

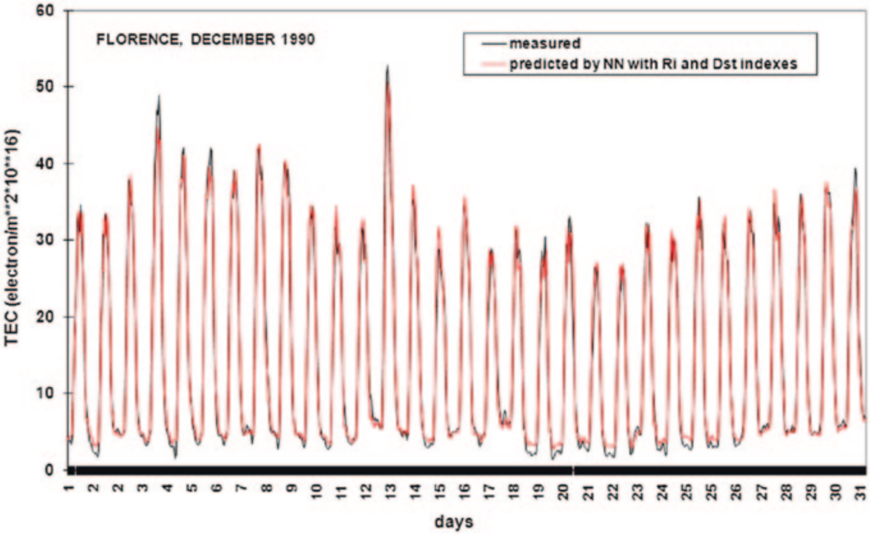


**Fig. 8.6** Measured and 1-hour ahead forecast *TEC* values for Florence (43.8° N, 11.2° E) in December 1990

The initial success of these investigations are self-evident. However, additional studies are recommended to researchers, data analysts and theoreticians to develop a new generation of ionospheric forecasting models based on artificial intelligence applications.

### 8.3.3 Forecasting Maps

Exploring the expansion of ionospheric forecasting from its inception, when it was little more than a curiosity of the scientific and amateur radio communities, to the present day when it is a recognized national and international service, inevitably leads to Short-Term Ionospheric Forecasting (STIF), a good example to follow at ionosonde stations where local and regional ionospheric modelling is required. STIF is an operational forecasting tool for the European region based on continuous monitoring of the ionosphere and developed at Rutherford Appleton Laboratory, Chilton, UK, which was one of the first space weather facilities for radio communications users available on the World Wide Web (http://www.rcru.rl.ac.uk/iono/maps.htm) for interactive use over a number of years.

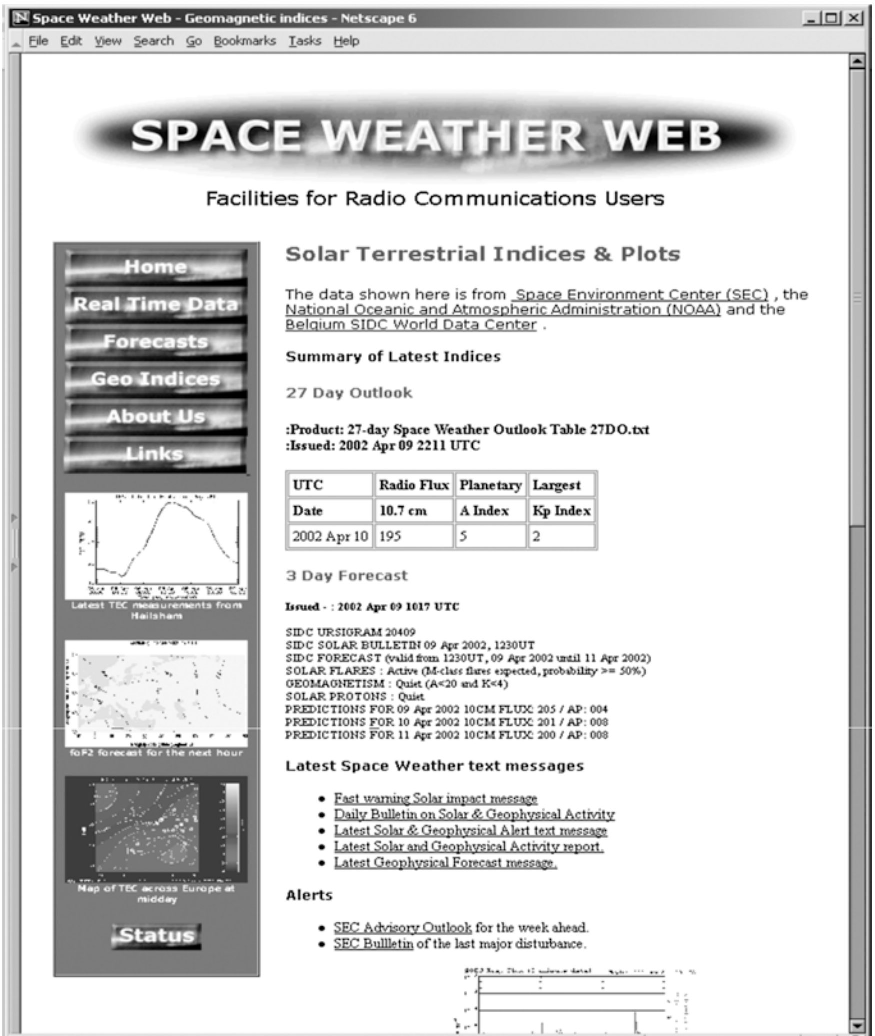
Figure 8.7 shows the front page of the portal with features that include among other products: (1) Near real-time *TEC* from 81 European IGS RINEX files; (2) Near real-time *foF2* and *MUF(3000)F2*; (3) Forecast maps of *foF2*, *MUF(3000)F2* and modelled *TEC* from 23 ionosonde stations; (4) Archive of all data and images; (5) Real-time solar-terrestrial indices; (6) Plots of monthly medians from selected stations. It is extremely important to ensure that these automatically running facilities are rigorously tested for the quality of the information provided.

Measurements of the critical frequency *foF2* and the Maximum Usable Frequency for a 3000 km range *MUF(3000)F2* are provided by 23 European and Asian ionosonde stations. These are sent by e-mail in the form of URSIGRAM messages. This network of up to 23 ground-based vertical incidence ionosonde stations, listed in Table 8.5, provides the basic inputs for the region of interest (10° W–90° E, 30°–70° N) with data updated every 24 h.

The autocorrelation procedure described in Sect. 8.3.1, developed for the short-term prediction of ionospheric characteristics, is applied to produce forecast values of *foF2* and *MUF(3000)F2* at integer hours UT up to 72 h ahead at each vertical incidence station where sufficient measurements are available. This permits users to always have a forecast available for up to 24 h ahead. Values from the past 60 days are used to construct the autoregressive filter. The characteristic *MUF(3000)F2* is derived from measured values of *foF2* and *M(3000)F2* according to the Eq. 3.3.9 in Sect. 3.3.4.

Contour maps are drawn using the Kriging interpolation technique, which is a gridding procedure using regional variable theory to calculate the autocorrelation between values whose weights depend upon their separations (see Chap. 6). The grid resolution is 2.5 °× 5 °. A linear variogram is used, together with an anisotropy factor which gives correspondingly greater weight to measurements along the longitudinal axis. Contour maps of forecast *foF2* and *MUF(3000)F2* values and of most recently available measurements are then produced. These are updated daily at fixed times.

Examples of *foF2* forecast and measurement maps are shown respectively in Figs. 8.8 and 8.9 for 31 and 29 January 2002 at 1700 UT. Similar examples for *MUF(3000)F2* are shown in Figs. 8.10 and 8.11, for the same dates respectively.



Solar Terrestrial Indices & Plots

The data shown here is from Space Environment Center (SEC) , the National Oceanic and Atmospheric Administration (NOAA) and the Belgium SIDC World Data Center .

Summary of Latest Indices

27 Day Outlook

:Product: 27-day Space Weather Outlook Table 27DO.txt
:Issued: 2002 Apr 09 2211 UTC

| UTC | Radio Flux | Planetary | Largest |
|---|---|---|---|
| Date | 10.7 cm | A Index | Kp Index |
| 2002 Apr 10 | 195 | 5 | 2 |

3 Day Forecast

Issued - : 2002 Apr 09 1017 UTC

SIDC URSIGRAM 20409
SIDC SOLAR BULLETIN 09 Apr 2002, 1230UT
SIDC FORECAST (valid from 1230UT, 09 Apr 2002 until 11 Apr 2002)
SOLAR FLARES : Active (M-class flares expected, probability >= 50%)
GEOMAGNETISM : Quiet (A<20 and K<4)
SOLAR PROTONS : Quiet
PREDICTIONS FOR 09 Apr 2002 10CM FLUX: 205 / AP: 004
PREDICTIONS FOR 10 Apr 2002 10CM FLUX: 201 / AP: 008
PREDICTIONS FOR 11 Apr 2002 10CM FLUX: 200 / AP: 008

Latest Space Weather text messages

- Fast warning Solar impact message
- Daily Bulletin on Solar & Geophysical Activity
- Latest Solar & Geophysical Alert text message
- Latest Solar and Geophysical Activity report.
- Latest Geophysical Forecast message.

Alerts

- SEC Advisory Outlook for the week ahead.
- SEC Bullletin of the last major disturbance.

**Fig. 8.7** Front page of the Space Weather Web portal for radio communication users

Comparisons were made between measured and forecast values, and the sensitivity of contour maps to errors was examined. Using the grid values, a statistical comparison was made between maps obtained from forecast and measured values of *foF2* and *MUF(3000)F2* for the same day at 0000, 0600, 1200, and 1800 UT on 8 widely separated days in March and June 1998. The forecast values selected were those deduced for one day ahead of the measured values. Overall, the *RMS* differences were 0.58 MHz for *foF2* and 2.14 MHz for *MUF(3000)F2*. This is to be expected, as *M(3000)F2* factors are typically about 3 and because there will

**Table 8.5** List of contributing vertical incidence (VI) ionosonde stations with their URSI codes and geographical coordinates (see red crosses in Figs. 8.9 and 8.11)

| VI Station | URSI code | Latitude (°N) | Longitude (°E) |
|---|---|---|---|
| Chilton | RL952 | 51.6 | 358.7 |
| Lannion | LN047 | 48.7 | 356.6 |
| Lerwick | LE061 | 60.1 | 358.4 |
| El Arenosillo | EA036 | 37.1 | 353.2 |
| Ebre | EB040 | 40.8 | 0.5 |
| Poitiers | PT046 | 46.6 | 0.3 |
| Rome | RO041 | 41.9 | 12.5 |
| Juliusruh | JR055 | 54.6 | 13.4 |
| Pruhonice | PQ052 | 50.0 | 14.6 |
| Warsaw | MZ152 | 52.2 | 21.2 |
| Lycksele | LY164 | 64.6 | 18.8 |
| Uppsala | UP158 | 59.8 | 17.6 |
| Kiruna | KI167 | 67.8 | 20.4 |
| Sofia | SQ143 | 42.7 | 23.4 |
| St. Petersburg | LD160 | 59.9 | 30.7 |
| Moscow | MO155 | 55.5 | 37.3 |
| Rostov | RV149 | 47.2 | 39.7 |
| Ashkhabad | AS237 | 37.9 | 58.3 |
| Salekhard | SD266 | 66.5 | 66.5 |
| Tashkent | TQ251 | 41.3 | 69.9 |
| Novosibirsk | NS355 | 55.0 | 82.0 |
| Tomsk | TK356 | 56.0 | 85.0 |
| Tunguska | TZ362 | 61.0 | 90.0 |

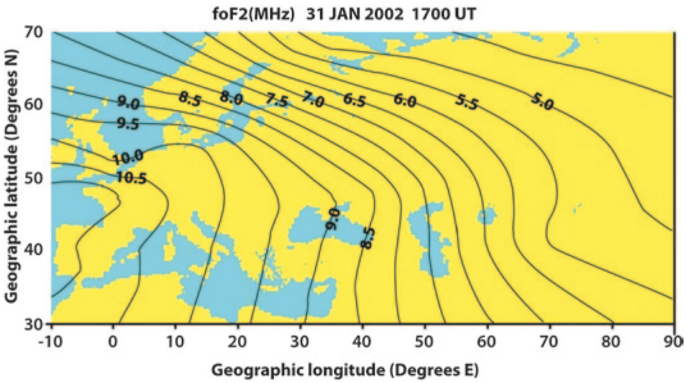


**Fig. 8.8** *foF2* map derived from forecast values for 24 h ahead for 31 January 2002 at 1700 UT

also be measurement errors of this characteristic. The overall normalised *RMS*, *NRMSE*, was 1.02, though on individual days this value was usually higher than this, reflecting the occasions in which the mean difference departed from zero.

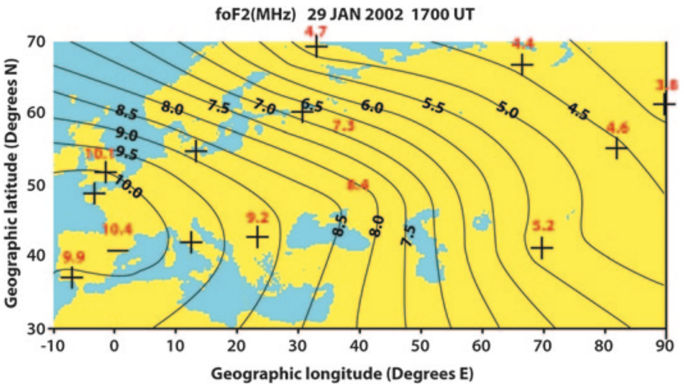


**Fig. 8.9** *foF2* map derived from measured values for 29 January 2002 at 1700 UT. The *crosses* indicate measured values at ionosonde stations

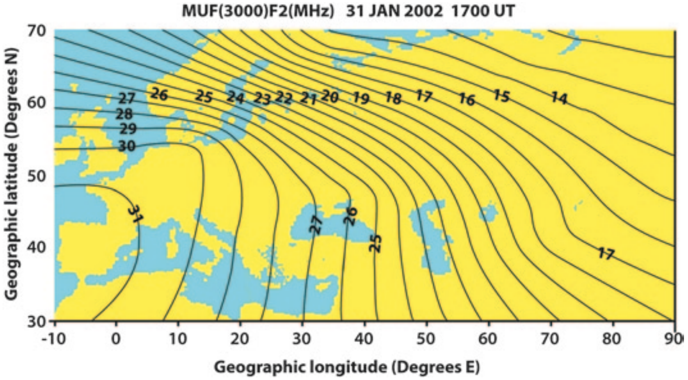


**Fig. 8.10** *MUF(3000)F2* map derived from forecast values for 24 h ahead for 31 January 2002 at 1700 UT

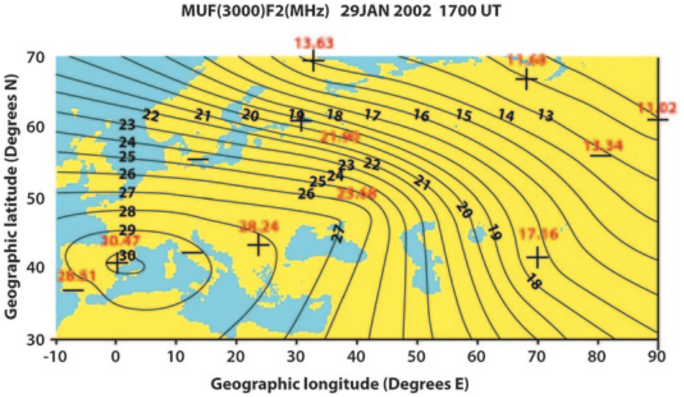


**Fig. 8.11** *MUF(3000)F2* map derived from measured values for 29 January 2002 at 1700 UT. The *crosses* indicate measured values at ionosonde stations

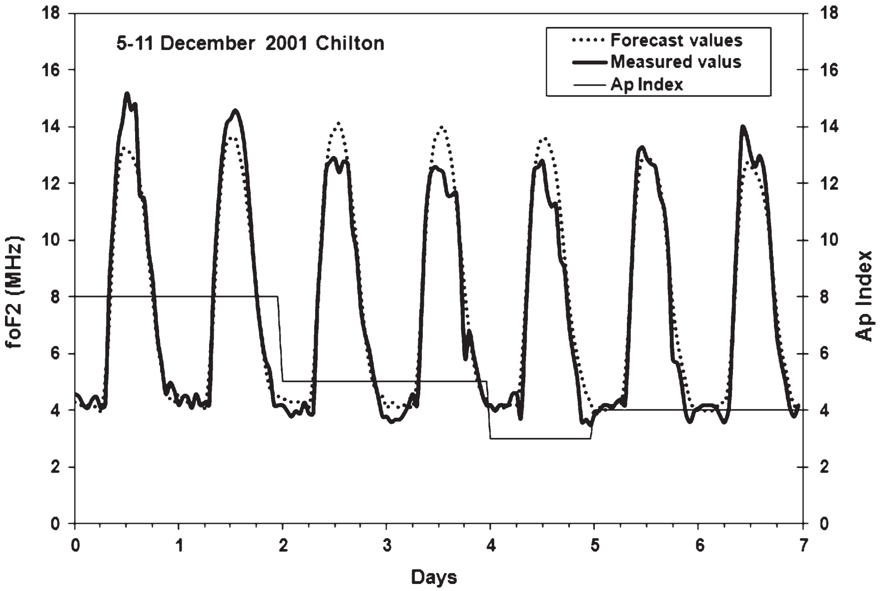


**Fig. 8.12** Typical STIF *foF2* results and measurements at the Chilton (51.6° N, 358.7° E) ionosonde station during quiet geomagnetic conditions with *Ap* values as an indicator of low geomagnetic activity ($RMSE = 0.76$ MHz; $NRMSE = 0.20$)

The STIF utility has been running continuously since the beginning of 1998, proving to be a useful tool to HF radio users who need up-to-date information on ionospheric conditions to meet their operational requirements. In addition, this facility provides a unique database for assessing the accuracy of forecasting when compared to individual measurements. Figs. 8.12 and 8.13 show typical comparisons between the *foF2* produced by the STIF programs and the data values measured by the Chilton ionosonde.

The periods shown are during the seven most geomagnetically quiet and most disturbed days in the months of December and April 2001 at the maximum of solar cycle 23. The *Ap* index shown in Fig. 8.12 indicates $Ap < 10$ during a geomagnetically quiet period. Under these conditions the forecasting technique performs excellently with a typical $RMSE = 0.76$ MHz and $NRMSE = 0.20$. The forecast values perform less well during extremely disturbed ionospheric conditions, as shown in Fig. 8.13 when the *Ap* index almost reaches 160. Here the $RMSE = 2.10$ MHz and $NRMSE = 0.80$. The forecast values do not immediately pick up the sudden change in the *Ap* index. There is a large discrepancy in the forecast and actual values for the day of the highest *Ap*. This can largely be attributed to the fact that the forecasting predictions are based on the previous day's data. However, the forecasting recovers quickly over the following days.

From an analysis of quiet and disturbed geomagnetic conditions it has been shown that the forecasting facility performs excellently during quiet periods and although it falters on extreme geomagnetic events, it still provides a valuable

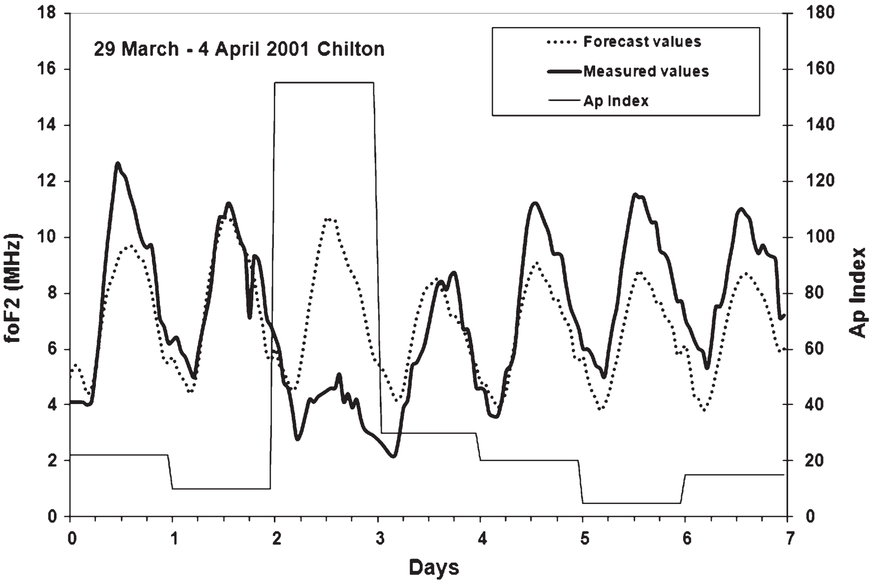


**Fig. 8.13** Typical STIF *foF2* results and measurements at the Chilton (51.6° N, 358.7° E) ionosonde station during disturbed geomagnetic conditions with *Ap* values as an indicator of high geomagnetic activity (*RMSE* = 2.10 MHz; *NRMSE* = 0.80)

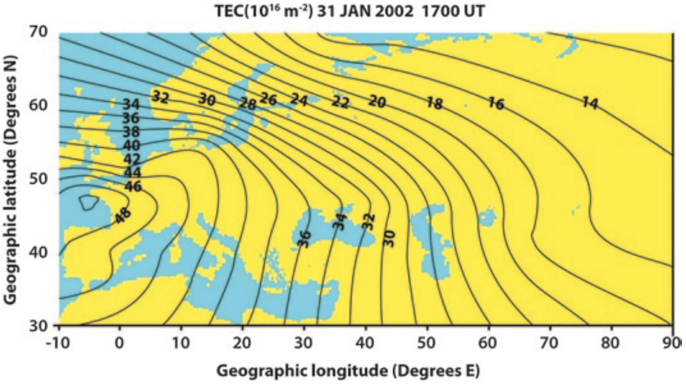


**Fig. 8.14** *TEC* map derived from 24 h ahead forecast values of *TEC* for 31 January 2002 at 1700 UT

predictor. As the STIF facility runs on predominately automatically scaled ionosonde data in order to provide a prompt service, the database also allows validation of the use of automatically scaled ionospheric parameters even under disturbed conditions.

An example of the *TEC* forecasting maps also produced from forecast values of *foF2* and *M(3000)F2,* indicated in Figs. 8.8 and 8.10, using the NeQuick ionospheric model (see Chap. 5) is shown in Fig. 8.14. A similar *TEC* archive map

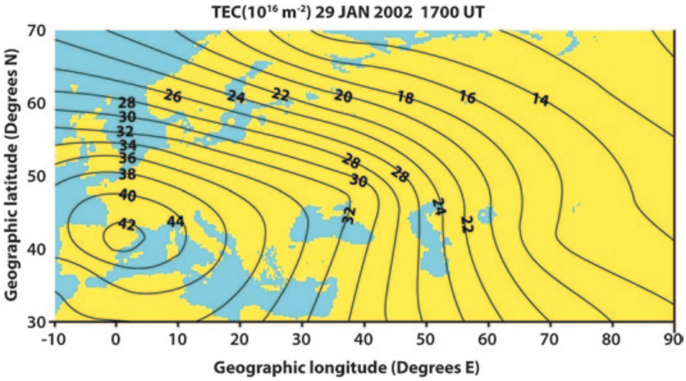


**Fig. 8.15** *TEC* map derived from values of *TEC* calculated from measured values for 29 January 2002 at 1700 UT

based on past measurements, indicated in Figs. 8.9 and 8.11, and the NeQuick ionospheric model is shown in Fig. 8.15.

In general, it is clear that accurate forecasting of ionospheric conditions and timely alert warnings is limited. Key factors posing the limitation are: (1) observational capability based on current density and frequency of ionospheric measurements; (2) present scientific understanding of the physics of the magnetosphere-ionosphere-thermosphere coupling mechanisms; and (3) sophistication and accuracy of ionospheric *N(h)* models and global maps of ionospheric characteristics such as *foF2*, *M(3000)F2*, and *VTEC*. In order to improve ionospheric forecasting capability, especially for tactical and satellite communications, it is necessary to significantly enhance the potential to frequently measure the ionsphere temporarily and spatially (see Chap. 10).

## Suggested Readings

Araujo-Pradere EA, Fuller-Rowell TJ, Codrescu MV (2002) STORM: An empirical storm-time ionospheric correction model 1. Model description. Radio Sci doi:10.1029/2001RS002467

Bamford RA, Stamper R, Cander LjR (2008) A comparison between the hourly autoscaled and manual scaled characteristics from the Chilton ionosonde from1996 to 2004. Radio Sci doi:10.1029/2005RS003401

Cander LjR (1993) On the global and regional behavior of the mid-latitude ionosphere. J Atmos Terr Phys 55:1543–1551

Cander LjR (2003) Towards forecasting and mapping ionospheric space weather under the COST actions. Adv Space Res 31:957–964

Cander LjR, Haralambous H (2011) On the importance of total electron content enhancements during the extreme solar minimum. Adv Space Res 47:304–311

Cander LjR, Milosaviljevic MG, Stankovic S, Tomasevic S (1998) Ionospheric forecasting technique by artificial neural network. Electron Lett 34:1573–1574

Ciraolo L, Spalla P (1994) An analysis of consistency of TEC evaluated using Pseudo-Range GPS observations. Proc. International Beacon Satellite Symposium, 21–24

Dick MI, Levy MF, Cander LjR, Kutiev I, Muhtarov P (1999) Short-term ionospheric forecasting over Europe. IEE Conference on Antennas and Propagation 461:105–107

Hargreaves JK (1979) The Upper Atmosphere and Solar-Terrestrial Relations. Van Nostrand Reinhold Co. Ltd., New York

Haykin S (1994) Neural networks—a comprehensive foundation. Macmillan College Publishing Company, New York

Korn GA, Korn TM (1968) Mathematical Handbook. McGraw-Hill Book, New York

Kutiev I, Muhtarov P, Cander LjR, Levy MF (1999) Short-term prediction of ionospheric parameters based on autocorrelation analysis. Ann Geofis 42:121–127

Lamming X, Cander LjR (1999) Monthly median foF2 modelling COST251 area by neural networks. Phys Chem Earth 24:349–354

McKinnell L-A, Friedrich M (2007) A neural network-based ionospheric model for the auroral zone. J Atmos Sol-Terr Phys 69:1459–1470

Mikhailov AV, Depuev VH, Depueva AH (2007) Short-term foF2 forecast: Present day state of art in space weather: Research towards applications in Europe. Astrophys Space Sci 344:169–184

Mitra AP (1974) Ionospheric effects of solar flares. D. Reidel Publishing Company Dordrecht-Holland, Boston

Muhtarov P, Kutiev I, Cander LjR (2002) Geomagnetically correlated autoregression model for short-term prediction of ionospheric parameters. Inverse Probl 18:49–65

Oyeyemi OE, McKinnell L-A (2008) A new global F2 peak electron density model for the International Reference Ionosphere (IRI). Adv Space Res 42:645–658

Prölss GW (1995) Ionospheric F-region storms. In: Volland H (ed) Handbook of Atmospheric Electrodynamics, CRC Press, Boca Raton

Stanislawska I, Juchnikowski G, Cander LjR (1996) Kriging method for instantaneous mapping at low and equatorial latitudes. Adv Space Res 18:217–220

Tsagouri I (2011) Evaluation of the performance of DIAS ionospheric forecasting models, J. Space Weather Space Clim. doi: 10.1051/swsc/2011110003

Tulunay E, Tulunay Y, Senalp ET, Cander LjR (2004) Forecasting GPS TEC using the neural network technique: A further demonstration. Bulg Geophys J 30:53–61

Ünal İ, Şenalp ET, Yeşil A, Tulunay E, Tulunay Y (2011) Performance of IRI-based ionospheric critical frequency calculations with reference to forecasting. Radio Sci doi:10.1029/2010RS004428

Williscroft L-A, Poole AWA (1996) Neural networks, foF2, sunspot number and magnetic activity. Geophys Res Lett 23:3659–3662

Wintoft P, Cander LjR (2000) Twenty-four hour predictions of foF2 using time delay neural networks. Radio Sci 35:395–408

# Chapter 9
# Prediction and Nowcasting for HF Applications and Radio Links

## 9.1 Introduction

As mentioned in the historical notes in the Introduction, the HF electromagnetic spectrum of 3–30 MHz became the principal waveband for long distance radio communications during the 20th century. Right from the beginning of ionospheric research it was clear that changes and intrinsic variability of the reflecting layers in the ionosphere are crucial for the performance of radio communications and when choosing the best radio frequency to use. Predicting the status of the ionosphere and assessing which parameters should be used for HF radio communication resulted in the development of geophysical and empirical models as well as statistical methods for collecting ionospheric measurements. In order to satisfy the needs of HF radio frequency planners, the prediction of hourly monthly medians for a given geophysical observation was implemented. Prediction for a given epoch of an hourly monthly median for a given ionospheric parameter, for example the critical frequency of the F region relative to the maximum reflectivity of the ionosphere, made it possible to establish that this value can vary by more than 50 % relative to the other values for the same time in the month.

The prediction of this value is extremely important for HF frequency planners who need to guarantee a radio link, but it should not be confused with the quiet condition of the ionosphere. Nevertheless, hourly monthly median values are often considered a close equivalent to a quiet ionosphere and they represent the background for day by day forecasting and real-time nowcasting.

Previous chapters described the solar origin of the terrestrial ionosphere, its vertical structure, and the global and regional morphology resulting from solar influence. After these fundamental observations, in order to predict the geophysical status of the ionosphere for a given epoch, it is necessary to define a basic

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_9,

procedure that establishes operating parameters, which effectively means the range of useful radio frequencies for a given radio link:

1. Assess the predicted solar activity index as an input for the geophysical model for the given epoch; depending on the model this could be the sunspot number $R_{12}$, the $T$ index, or the solar flux $\Phi$;
2. Consider the geometrical aspect of the radio link between the transmitting point and receiving point, including very short, near vertical links and very long links with multiple reflections from the Earth and ionosphere (Fig. 9.1);
3. Predict for a given epoch the two ionospheric characteristics *foF*2 and *M*(3000)*F*2, useful for calculating the operating parameters, at the points where reflection occurs in the ionosphere using the global or regional maps (Fig. 9.2);
4. In the case of nowcasting, correct for "climatological" behaviour based on the hourly monthly median model with real-time measurements as new input;
5. Calculate the operating parameters taking into account the ionospheric conditions as well as the characteristics of the radio equipment.

The basic principles of how to calculate the operating parameters *MUF* (Maximum Usable Frequency) and *LUF* (Lowest Usable Frequency) are reported in this Chapter as well as typical outputs and results, including hourly diagrams for given and standard distances, and hourly HF maps of skip distance, showing isolines of equal *MUF*.

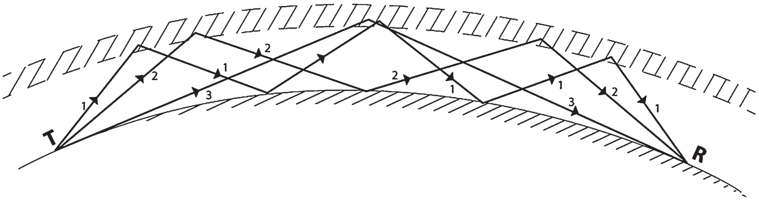


**Fig. 9.1** Simple scheme of a radio link: a signal from a transmitting point T to a receiving point R can be received with only one hop and one reflection in the ionosphere (ray 3), two hops (ray 2), and even 3 hops (ray 1)

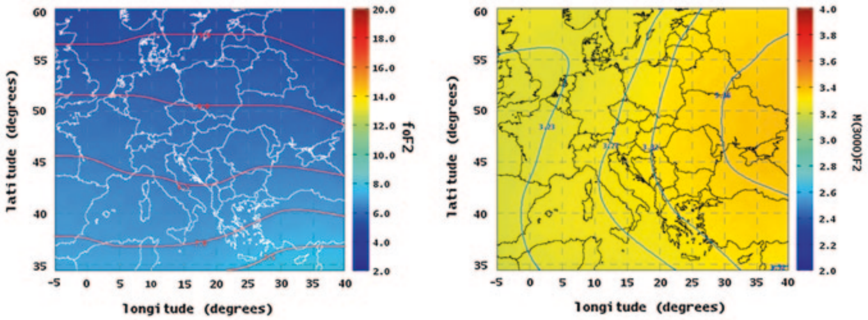


**Fig. 9.2** An example of regional maps of the predicted ionospheric characteristics *foF*2 and *M*(3000)*F*2

## 9.2 HF Ionospheric Performance Predictions

### 9.2.1 *Geometrical Aspects or Basic Principles of an HF Radio Link: The Secant Law and Martyn Theorem*

Figure 9.3 shows a simple scheme of an HF radio link demonstrating that the behaviour of different electromagnetic rays depends on the angle of elevation $\beta$. In Fig. 9.3 rays *a* and *b*, with a small angle of incidence $\varphi$ to the ionosphere, escape into space, rays *c*, *d*, *e*, and *f*, with an angle of incidence greater or equal than $\varphi_o$, are reflected by the ionosphere.

The critical angle of incidence is defined by:

$$\varphi_\mathrm{o} = \arccos f_0/f \tag{9.2.1}$$

where $f_0$ is maximum frequency reflected at a vertical incidence, this being the critical frequency of the ionospheric layer, and $f$ the frequency of the rays. Rays reflected by the ionosphere will arrive at greater distances until the elevation angle $\beta$ is tangential to the ground. The skip distance is defined as the minimum distance that a ray coming back from the ionosphere is reflected. Within this distance, also known as the silent distance, only ground wave propagation is possible. This form of propagation for HF waves is very limited and is highly dependent on the altitude of the transmitting antenna and the composition of the reflecting points on the Earth, including sea, ice or rocks.

In a simplified model of a flat ionosphere over a flat Earth, as illustrated in Fig. 9.4, the reflection starts at point $\boldsymbol{\delta}$, at the beginning of the ionosphere, changes to downwards at point B, and achieves complete reflection at point $\boldsymbol{\varepsilon}$.

According to the theorem of Breit and Tuve in 1926 the time of propagation of the ray between the points TBR in Fig. 9.4 is equal to the time taken by light to cover the path through TCR in a vacuum. On the other hand Martyn's theorem in 1935 says that the virtual height $h'_{ob}$ of an oblique reflection at a given radio frequency $f$ is equal to the virtual height $h'_v$ corresponding to a vertical reflection of a frequency $f_v$ so:

$$f_v = f \cos \varphi_o \tag{9.2.2}$$

According to the simple scheme in Fig. 9.4, and temporarily ignoring other complex behaviours, it could be considered that for any oblique radio frequency $f_{ob}$, reflection will occur at the vertical height $h'_v = \mathrm{CA}$. The equation:

$$f_{ob} = f_v \sec \varphi_o \tag{9.2.3}$$

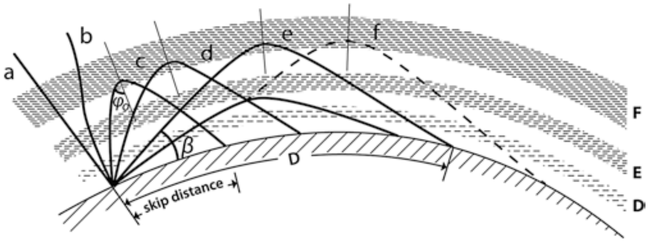
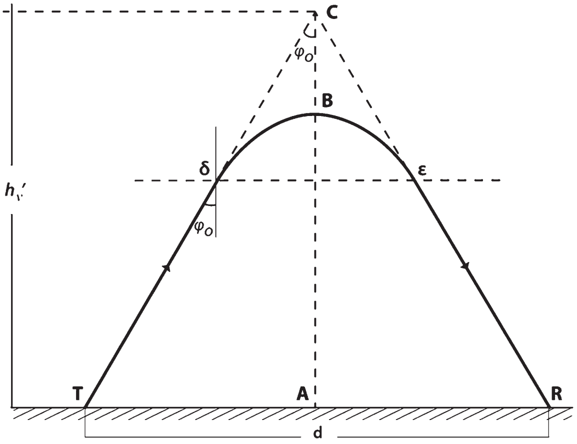


**Fig. 9.3** Simple scheme of a radio link

**Fig. 9.4** Simplified model of a flat ionosphere over a flat Earth



is known as the secant law. Referring to the same simple figure with the initial approximation of a flat Earth and flat ionosphere,

$$\sec\varphi_o = \sqrt{1 + (d/2h_v')}^2 \tag{9.2.4}$$

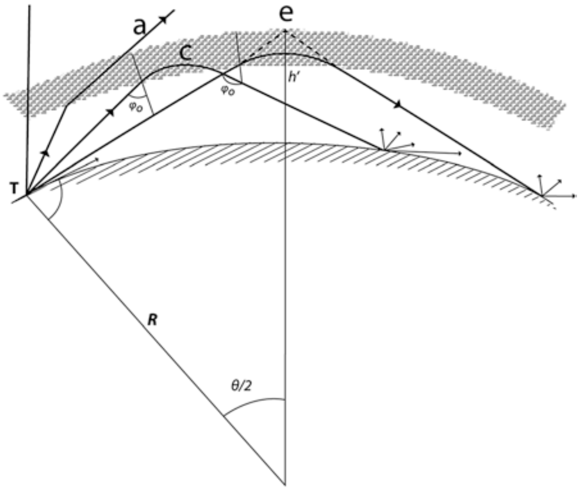
where $d$ is the distance between the transmitter T and the receiver R for flat condition.

With a symmetrical and spherical ionosphere over a spherical Earth, as in Fig. 9.5, the critical angle of incidence is defined by:

$$\varphi_0 = \arctan\frac{\sin\frac{\theta}{2}}{1 + \frac{h'}{R} - \cos\frac{\theta}{2}} \tag{9.2.5}$$

where $\theta$, $h'$ and $R$ are given in Fig. 9.5.

**Fig. 9.5** Scheme of different radio links on a real spherical Earth and symmetrical spherical ionosphere with rays a, c, and e as in Fig. 9.3



### 9.2.2 MUF Definition and Calculation

Taking into account Eq. 9.2.3, as a first approximation, if the vertical frequency $f_v$ is the critical frequency $f_o$ defined in Chap. 3, the maximum radio frequency reflected along an oblique path with only one reflection is defined as the Maximum Usable Frequency (*MUF*):

$$MUF\ (D) = f_o \sec \varphi_o \tag{9.2.6}$$

where $\sec \varphi_o$ comes from Eq. 9.2.5 which depends on the geometry of the link and consequently on the distance $D$ between the transmitter and the receiver. More accurately the real *MUF* is given by the maximum value of the product $f_v \sec \varphi_o$. In Eq. 9.2.5 it is clear that MUF increases with distance $D$ up to 4,000 km, which can be considered the upper limit for a single hop when the elevation angle of the ray is almost tangential to the Earth.

It is important to note that the *MUF* for a given distance $D$ is dependent only on the maximum vertical frequency reflected at the mid point, $f_o$, and on the geometry of the radio link. The characteristics of the radio hardware, including power of transmission or antenna system, do not influence the *MUF* at a given distance. Similarly, the minimum distance of reflection or skip distance depends on the reflectivity of the ionosphere at the mid point of distance $D$.

The *MUF* can be calculated directly from the ionograms obtained during routine vertical incidence soundings. Using the two Eqs. 9.2.3 and 9.2.5 it is possible for a given oblique frequency $f_{ob}$ to draw a curve as a function of $h'$ and $f_v$ known as the transmission curve, introduced for the first time by N. Smith in 1939. The intersections of this curve with the curve of the function $h' = \mathrm{f}\ (f_v)$ of the ionogram, gives graphical solutions to the two equations, representing the two virtual heights of reflection of a transmission with two propagation paths, points a and a′ in Fig. 9.6.

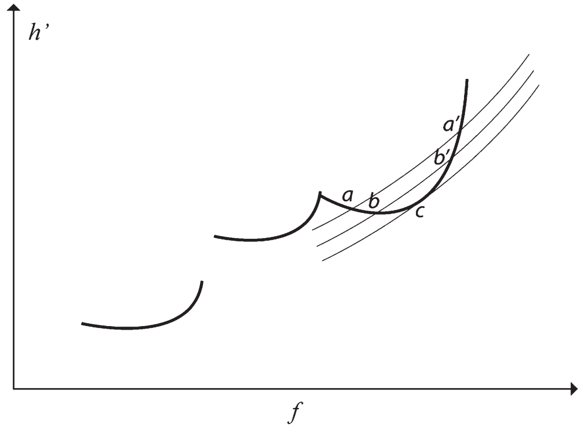


**Fig. 9.6** Family of curves given by the equation $f_{ob} = f_v \sec \varphi_o$, which intersect the curve of the ionogram produced by the equation $h' = f\,(f_v)$

Next, by increasing the given oblique frequency a family of transmission curves can be obtained, which intersect the ionogram curve at increasingly close points b and b′, until they become tangential as in point c. At this point only one propagation path is possible and the oblique frequency, corresponding to the tangential transmission curve, is the maximum usable frequency for that distance. In the past the transmission curves of the standard distance 3,000 km were drawn on a transparent slide, obviously maintaining the same scale of frequency and virtual height, as in Fig. 9.7. Laying it on top of an ionogram it is possible to read the *MUF*(3000)F2 directly.

It is important to add that both the ionosphere and the Earth are curved and the virtual height also depends on the electron density profile. For practical purposes it is sufficient to correct the secant law with:

$$f_{ob} = k f_v \sec \varphi_o \tag{9.2.7}$$

where $k$ is a correction factor and a function of distance and the real height of reflection, which varies from 1 to 1.2.

The $M$ factor, introduced in Chap. 3 as one of the parameters routinely scaled and mapped, is then defined as:

$$M(3000)F2 = MUF(3000)F2/f_oF2 \tag{9.2.8}$$

and used in practice. According to Eq. 3.2.73 in Chap. 3, for vertical propagation with $D = 0$ the $MUF(0)$ *is*:

$$MUF(0) = f_oF2 + f_H/2 \tag{9.2.9}$$

where $f_oF2 + f_H/2$ is the frequency of the reflected extraordinary ray. Today the $M(3000)F2$ factor is computed automatically using autoscaling methods as well as the $M(D)$ of a given distance $D$ instead of the graphical methods or nomograms used in the remote past. Among HF users different definitions of *MUF* were established. To resolve this the International Radio Consultative Committee (CCIR) proposed:

(a) Basic *MUF* is the highest frequency of propagation between two points of a radio link established only by ionospheric reflection and permitted by the geophysical conditions along the radio path.

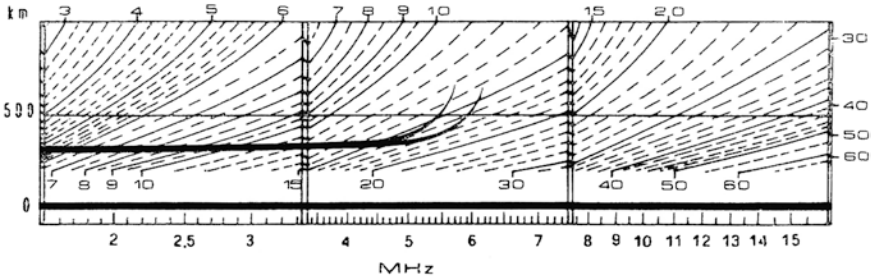


**Fig. 9.7** Example of a transparent slide on which a family of curves are drawn for the standard distance of 3,000 km and superimposed on a given ionogram. The *MUF*(3000)*F*2 in this case is 16 MHz, represented by the tangential curve to the ordinary trace

(b) Operating *MUF* is the highest frequency that permits a radio link between two points not only under ionospheric and geophysical conditions but also under given working conditions, like class of service, or signal-to-noise ratio, and including the hardware specifications like transmission power and gain of the antenna used.

The $M(3000)F1$ is also obtained in the same way described above and is routinely scaled at vertical ionospheric stations.

The *MUF* over the E region can be calculated using the model for $f_oE$ in Chap. 4, applying a standard height of reflection of 100–110 km in the geometry of the link to calculate $\sec\varphi_E$. This prediction is important to evaluate the cut-off frequency produced by the underlying region, or in the case of radio links extending West to East when a portion is under diurnal conditions with a strong E region, and another portion is under nocturnal conditions, with a predominant F region.

### 9.2.3 Determination of the Path of an Electromagnetic Wave in the Ionosphere

Propagation and reflection of an electromagnetic wave in the ionosphere is such a complex phenomenon that the simple schemes shown previously in Figs. 9.1, 9.3, 9.4, and 9.5 modelled using the secant law and Breit and Tuve's, and Martyn's theorems are not adequate to completely represent the real path of rays.

For example, in Fig. 9.8 the possible ray paths with a fixed frequency and an increased elevation angle are traced applying the index of refraction calculated using the S. Chapman model. Note how a particular ray, with an elevation angle slightly greater than the angle related to the skip distance, does not propagate in free space but travels in the ionosphere, returning to the ground over longer distances. This ray, known as a Pedersen ray, is reflected and propagates at higher altitudes where the plasma frequency is lower than the maximum electron density.

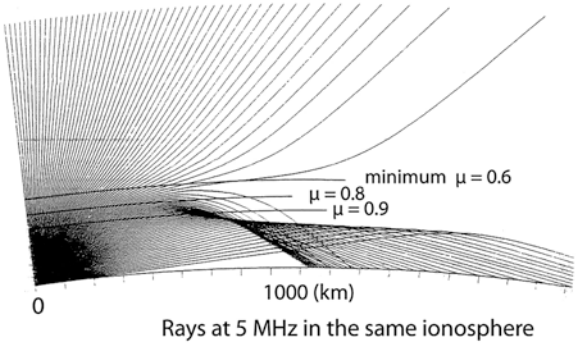


**Fig. 9.8** Possible ray paths of a fixed frequency with increasing elevation angle (from Bradley 1989)

A ray tracing computational technique is necessary in order to solve the problem of precisely determining the arrival point of a radio signal after travelling through the ionosphere.

The basic principle of the ray tracing technique is to apply the ray theory over a series of thin homogenous layers of the ionized medium and then to integrate the different contributions over the entire path. Figure 9.9 shows a scheme of concentric thin layers, each with an index of refraction $\mu$.

To determine the real distance $D$ between the transmitting point A and the arrival point B of Fig. 9.10, given by $D = R\theta$, it is necessary to calculate $\theta$, the angle at the Earth's centre, already knowing its radius $R$.

Introducing the Bougeur's law that generalizes the Snell law for media with spherical symmetry:

$$\mu H \sin\varphi = \mu_0 H_0 \sin\varphi_0 \tag{9.2.10}$$

with $\mu$ the index of refraction at altitude $H = (R + h)$, $\varphi$ the incidence angle of the ray at any point of the ionosphere, $\mu_0 = 1$ the index of refraction, and $\varphi_0$ the angle of incidence at altitude $H_0 = (R + h_0)$ at the bottom of the ionosphere.

By geometry of the ray path (details in Fig. 9.10):

$$(R + h)d\theta = \tan\varphi \cdot dh \tag{9.2.11}$$

inserting Snell's law:

$$\sin\varphi = \frac{(R + h_0)\sin\varphi_0}{\mu(R + h)} \tag{9.2.12}$$

and recalling the fundamental law of trigonometry:

$$\cos^2\varphi = 1 - \sin^2\varphi = 1 - \left[\frac{(R + h_0)\sin\varphi_0}{\mu(R + h)}\right]^2 \tag{9.2.13}$$

it follows that:

$$d\theta = \frac{\tan\varphi}{(R + h)}dh = \frac{dh}{(R + h)}\frac{(R + h_0)\sin\varphi_0 \big/ \mu(R + h)}{\sqrt{1 - [(R + h_0)\sin\varphi_0 \big/ \mu(R + h)]^2}} \tag{9.2.14}$$

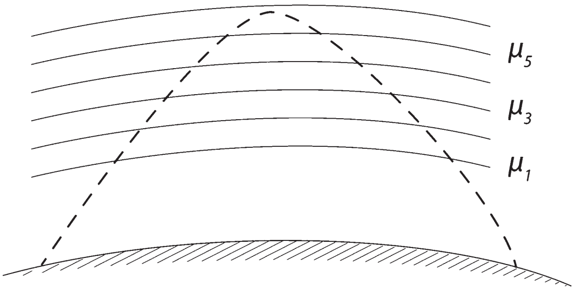
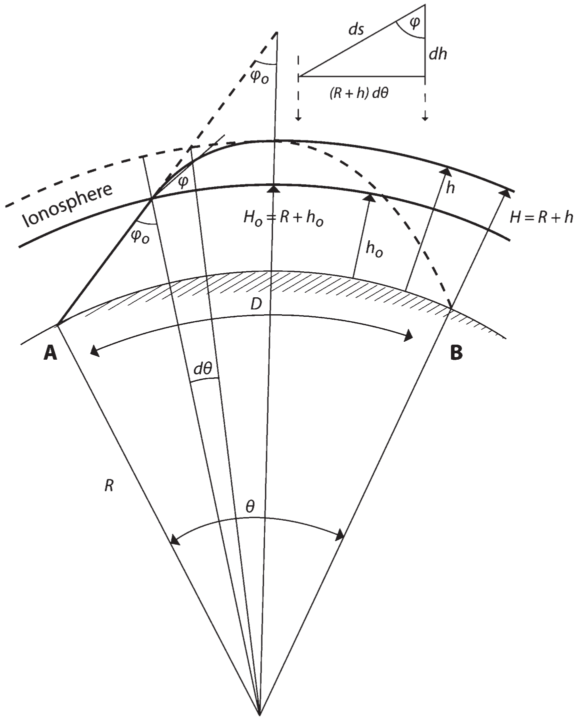


**Fig. 9.9** Scheme of concentric thin layers of the ionized medium, each with an index of refraction $\mu$

**Fig. 9.10** Geometry of the ray path between points A and B at a distance *D*



Hence, the angle $\theta$ can be obtained by integrating from $h_0$ to $h$:

$$\theta = 2\int_{h_0}^{h} d\theta = 2\int_{h_0}^{h} \frac{dh}{R+h} \frac{(R+h_0)\sin\varphi_0 \big/ \mu(R+h)}{\sqrt{1-[(R+h_0)\sin\varphi_0 \big/ \mu(R+h)]^2}} \qquad (9.2.15)$$

Now the difficulty of this integral concerns the calculation of $\mu$ at every point of the path that may be derived by an analytical model of $N(h)$ profile, so that the equation can have an analytical solution, or by fitting analytical $N(h)$ profiles in many segments, by tabulating values of the electron density *versus* the altitude. In this elaborated computational process the deviation introduced by the magnetic field should also be considered. For an exhaustive treatment of ray tracing in the ionosphere, see K. G. Budden in 1961; for short but rigorous presentations see K. Davies in 1990, L. F. McNamara in 1991, and P. A. Bradley in 1989.

### *9.2.4 Definition and Calculation of Attenuation and LUF*

While *MUF* is the maximum usable frequency in a given radio link, *LUF* is the lowest usable frequency in the same radio link. *MUF* is only dependent on the reflectivity of the ionosphere and the distance of the two terminals. *LUF* depends not only on the conditions of the ionospheric plasma and the geometry of the radio link, but also on the characteristics of the radio equipment used to establish the radio link.

This parameter is calculated by applying empirical models resulting from field measurement campaigns and it differs between the different national services. The basic physical principles on which they are all based are described below.

Generally the *LUF* is determined by the maximum absorption or attenuation of the electromagnetic signal between the transmitting equipment and the receiving system, taking into account its path through the ionosphere, the neutral atmosphere, and possible reflections from the ground and sea for multiple paths.

To calculate this parameter it is necessary to evaluate the total attenuation given by the sum of non ionospheric attenuation, which includes geometrical attenuation $L_g$, deviative $L_d$ attenuation due to the portion of the radio path inside ionospheric plasma and close to the point of reflection, and $L_n$ attenuation in which the electromagnetic wave does not undergo any deviation, mainly in the D region.

A first contribution to attenuation is obviously geometrical attenuation $Lg$ measured in dB (for the logarithm formalism see Appendix A), which clearly depends on the loss of energy due to the distribution of energy over the spherical surface:

$$[L_g]_{dB} = 20 \log S \tag{9.2.16}$$

where $S$ is the complete path length covered by the ray along the distance $D$ in km, the maximum circular arc over the Earth, and all the possible paths taking into account the height $h_s$ of the reflections. The distance $D$ between two points T and R over the Earth's surface with $\varphi_T$ and $\lambda_T$ and $\varphi_R$ and $\lambda_R$ their longitudes and latitudes, as shown in Fig. 9.11, is:

$$D = \vartheta \cdot \bar{R} \tag{9.2.17}$$

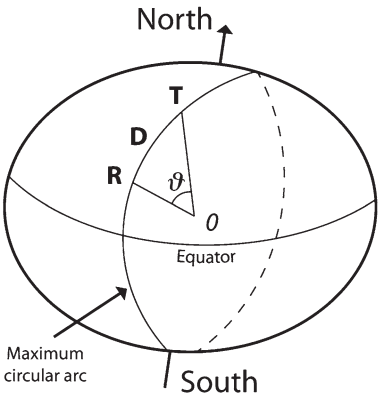


**Fig. 9.11** The maximum circular arc over the Earth between points T and R is given by the distance *D*

with:

$$\vartheta = \arccos\left[\sin\lambda_T \cdot \sin\lambda_R + \cos\lambda_T \cdot \cos\lambda_R \cdot \cos(\varphi_T - \varphi_R)\right] \quad (9.2.18)$$

the angle to the centre of the Earth, and $\bar{R}$ the average of terrestrial *radii* equal to 6,370 km.

Attenuation produced by multiple reflections from the ground or sea must also be added, depending on their composition or physical state, ice or fluid, and also depending on the radio frequency. Radio planning services solved this problem by producing an atlas and other graphical media that reflect empirical models based on "*una tantum*" measurement campaigns. However, it is clear that the longer the radio path, the higher the geometrical attenuation [$L_g$]. See for example the notes by France Telecom at http://www.iono.enst-bretagne.fr.

The assessment of deviative attenuation [$L_d$] is the result of a complex mathematical model that takes into account the incident radio frequency $f$, the maximum collision frequency $\upsilon_M$, the critical frequency $f_o$, the angle of incidence $\varphi_0$, and the semi thickness of the layer $y_m$:

$$[L_d]_{dB} = c\,\upsilon_M\,(y_m/f_o)\,(f/\sec^3\varphi_o) \quad (9.2.19)$$

with $c$ a constant.

Deviative attenuation is greater than non-deviative only when the frequency $f$ is close to the *MUF*. However, the result of this is an approximate relation, and when typical values for the E region and F region are applied it is so small that it is not usually considered, except for an additional contribution of a few dB.

Non-deviative attenuation [$L_n$] is mainly due to an oblique path through the D region and partially through the E region, so it depends on the condition of these diurnal layers:

$$[L_n]_{dB} = P(a + \cos\chi)^{3/4}\sec\alpha/(f + f_h)^2 \quad (9.2.20)$$

where $a$ is a constant, and $P$ is depending on the number of particles in the ionosphere and so proportional to the frequency of collisions $\upsilon$ and the electron density $N$, which in the ionospheric plasma are dependent on the height $h$ and on the solar activity index, $\alpha$ is the angle of penetration in the D region, $\chi$ the solar zenith angle, and $f$ and $f_h$ are the radio frequency and the gyrofrequency. This means that attenuation is strongly dependent on the time of day $\chi$, on the presence of the D region, where $N$ and $\upsilon$ are higher, and attenuation is inversely proportional to the square of the radio frequency $f$. The D region is the most absorbent region and is responsible for the marked attenuation of MF radio stations during the day, while these are instead received clearly during the night.

Now, to calculate the lowest practically usable frequency it is necessary to introduce the concept of maximum attenuation [$L_{max}$] or transmitting power [$P_T$]. It is given by the ratio of the power $P_{TX}$, or the intensity of electric field $E_{TX}$, measured at the transmitter antenna, and the power $P_{RX}$ or intensity $E_{RX}$ measured at the receiving antenna:

$$[P_T] = 10 \log \frac{P_{TX}}{P_{RX}} dB \tag{9.2.21}$$

Taking into account that $P_{TX}/P_{RX} = (E_{TX}/E_{RX})^2$ then:

$$[P_T] = 20 \log \frac{E_{TX}}{E_{RX}} dB \tag{9.2.22}$$

More precisely $P_{TX}$, is the power and $E_{TX}$, the intensity of the electric field emitted by the transmitter and measured at a distance of 1 km. $P_{RX}$ and $E_{RX}$ are instead the minimum power and minimum electric field required to establish a radio link under given characteristics of a service that needs to be maintained.

By using the logarithmic formalism Eq. 9.2.22 this can be expressed as:

$$[P_T] = [E_{TX}] - [E_{RX}]\, dB \tag{9.2.23}$$

To establish a radio link it is necessary that the $[P_T]$ given by Eq. 9.2.23 must be greater than the total attenuation:

$$[P_T] > [L_{max}] = [L_d] + [L_n] + \left[L_g\right] dB \tag{9.2.24}$$

Taking into account Eq. 9.2.20 and ignoring the deviative attenuation $[L_d]$, the minimum frequency for the diurnal *LUF* can be obtained for only one hop in the ionosphere:

$$(f + f_h)^2 \geq B \sec \alpha / (P_T - L_g) \tag{9.2.25}$$

with $B = P(a + \cos\chi)^{3/4}$ as defined above. Therefore for only one hop:

$$LUF_{day} = (B \sec \alpha / (P_T - L_g))^{1/2} - f_h \tag{9.2.26}$$

For multiple hops the different position of the Sun at all the points of reflection must be considered, and different attenuations must be added for each separate hop.

For the nocturnal *LUF* this simple relation is valid:

$$LUF_{night} = (([Lg] - [P_T] + c)/b) \tag{9.2.27}$$

where $b$ and $c$ are two empirical constants.

It is important also to consider the so called blanketing or cut-off frequency, this being the case in which an electromagnetic wave is reflected first by the E region and so preventing propagation up into the F region.

This frequency, defined as $LUF_{bl}$, is related to the angle of incidence $\varphi_E$ over the E region (see Fig. 9.3) and the critical frequency *foE* of the E region according to the relation:

$$LUF_{bl} = f_o E \sec \varphi_E \tag{9.2.28}$$

$LUF_{bl}$ is equivalent to the maximum oblique frequency over the E region, or in other words the *MUF* over the same region.

### 9.2.4.1 How to Calculate the Transmitting Power

As an example, in order to calculate the transmitting power $P_T$ defined in Eq. 9.2.21, the density of power $P$ at a distance $r$ is introduced:

$$P = \frac{W_T G_T}{4\pi r^2} \tag{9.2.29}$$

with $W_T$ the power applied to the transmitting antenna, $G_T$ the gain of the antenna system, and $4\pi r^2$ the spherical surface over which the power is distributed. $E$ is the intensity of the electric field at the point in question, and is related to the density power $P$ in the equation:

$$P = \frac{E^2}{Z_0} \tag{9.2.30}$$

where $Z_0 = 377$ Ohm is the impedance of vacuum. Then:

$$E = (377/4\pi r^2)^{1/2} (W_T G_T)^{1/2} \tag{9.2.31}$$

Assuming that in this example $r = 1$km; $W_T = 1$ kW, and the reference $G_T$ in comparison with an isotropic radiator is equal to 1.64, then $E_{TX}$ in μV/m, and $P_{TX}$ in kW are given by:

$$E_{TX} = 10^6 \left( \frac{377 \cdot 1{,}000 \cdot 1.64}{4\pi \cdot 10^6} \right)^{\frac{1}{2}} \cdot (P_{TX} G_T)^{\frac{1}{2}} \left[ \frac{\mu \text{V}}{\text{m}} \right] \tag{9.2.32}$$

and:

$$E_{TX} = 222{,}000 \cdot (P_{TX} G_T)^{\frac{1}{2}} \left[ \frac{\mu V}{\text{m}} \right] \quad \text{or:} \tag{9.2.33}$$

$$E_{TX}^2 = 222{,}000^2 \cdot (P_{TX} G_T) \tag{9.2.34}$$

which becomes in decibel units:

$$10 \log E_{TX}^2 = 10 \log 222{,}000^2 + 10 \log P_{TX} + 10 \log G_T \tag{9.2.35}$$

$$[E_{TX}] = 107 + [P_{TX}] + [G_T] \tag{9.2.36}$$

Therefore Eq. 9.2.36 is the relation that gives the intensity of the electric field $E_{TX}$ expressed in dB over μV/m at a distance of 1 km, with a gain referred to a dipole of λ/2 and when the power applied $P_{TX}$ is given in kW.

Now, to complete the calculation of the $P_T$ parameter, some brief considerations are useful regarding the value of $E_{RX}$, which can be calculated with the relation:

$$[E_{RX}] = 20\log(E_{Rm}/\mu\mathrm{V/m}) \tag{9.2.37}$$

where $E_{Rm}$ is the minimum intensity of the electron field sufficient to maintain the radio link under service requirements. This parameter can also be obtained by:

$$[E_{Rm}] = [C] + [S] \tag{9.2.38}$$

as the result of two components: the radio noise around the receiving antenna $C$ and the quality of the service required, $S$. The radio noise component, this being the minimum intensity required to maintain the radio link, depends on the time, season, and also location because receivers are strongly influenced by industrial radio noise, radio traffic, and the presence of radio broadcasting emissions. The $S$ factor depends on the required quality of service. For telegraphy radio links with only Morse code the correction factor $S$ can be approximated by less than 3 dB. This correction factor should be higher for telephony transmissions and much higher still for a broadcasting agency that needs top reception quality adequate for a symphony.

### 9.2.5 Point-to-Point HF Prediction and Nowcasting

All *MUF* calculations are valid both for long-term predictions, usually monthly medians, and for nowcasting. What differs is obviously the prediction or the nowcasting of the *foF2* and $M(3000)F2$ at all points where there is reflection in the ionosphere. The *LUF*, due to the fact that this parameter is derived not only from the characteristics of the radio equipment but also from empirical models and constants, is not variable day by day but only hourly, monthly, and seasonally at a given location.

Using the mapping procedures described in Chap. 6 it is possible to predict the values of $foF2$ and $M(3000)F2$ at any point on the globe or for a region during a given epoch, month and hour. If the mapping procedure is a long-term prediction of ionospheric characteristics based on monthly median statistics, the prediction will give monthly median values of $foF2$ and $M(3000)F2$ for each hour and then the monthly median *MUF*. Otherwise, if the mapping procedure is a forecast or a nowcast then short-term forecasting or even real-time evaluation of the *MUF* is obtained.

It is very important to note that the choice to establish a monthly median statistics and consequently the prediction of monthly median parameters was traditionally derived and codified in international rules and conventions, in response to the need by radio frequency planners for certainty that a given predicted frequency will be below the real *MUF* for at least 50 % of the days in a given month. This means that the monthly median *foF*2 and the monthly median *M*(3000)*F*2 used to calculate the *MUF* is less than 50 % of the daily *foF*2 and *M*(3000)*F*2 at the same hour in the month.

In the case of a simple radio link with only one reflection in the ionosphere, using the Eqs. 9.2.5 and 9.2.6 it is possible to build a graph for two given points of hourly behaviour of the *MUF*, and using the equations in Sect. 9.2.3 for the hourly behaviour of the *LUF* (Fig. 9.12).

The hourly diagram of *MUF* and *LUF* represents the range of frequencies that can be used for a given two points and the same radio equipment characteristics during all the days of the month, considering the monthly median statistics. In cases of radio links with more than one reflection on the ionosphere and on the Earth's surface, or even with reflections from both the E and F regions which are typical for long distance connections, the *MUF* and *LUF* graphs are obtained by combining the single curves, using the lowest values of *MUF* and highest values of *LUF*. Even if theoretically all the radio frequencies within this range could assure a connection, the attenuation is different for every frequency and it is minimum for one, defined as the optimum frequency. The optimum working frequency or *FOT* is the best working frequency across a given radio link and it is particularly useful when it is not possible to calculate the value of *LUF*. On the basis of experience, the frequency with the best signal-to-noise ratio is generally calculated as:

$$FOT = 0.85\,MUF \tag{9.2.39}$$

only if it is greater than *LUF*.

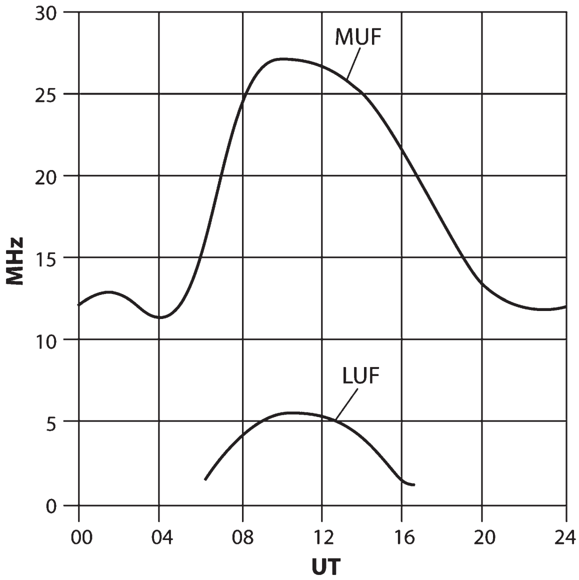


**Fig. 9.12** Example of hourly monthly median prediction of *MUF* and *LUF* for a generic distance and month

If the applied statistics take into account the lower quartile LQ and the upper quartile UQ of a set of hourly data in a month, as routinely reported in the ionospheric bulletins according to URSI conventions, it is possible to generate graphs of *MUF* and *LUF* with different levels of probability. The two graphs for the *MUF* and the two graphs for the *LUF* in Fig. 9.13 represent the hourly behaviour for two probabilities: 90 % and 30 %. This means that the *MUF* values are equal to or greater than the corresponding 90 % line, and the *LUF* values are equal to or lower than the corresponding 90 % line for 27 days in the month. In the case of 30 %, the *MUF* values are equal to or greater than the corresponding line and the *LUF* values are equal to or lower than the corresponding line for 10 days in the month.

Increasing the distance between the two points of the radio link will increase the *MUF* and *LUF* graphs. Considering that *MUF* is dependent on the geometry of the radio link, the reflectivity of the ionosphere, and the electron density, it cannot be arbitrarily changed. Therefore users must adapt their frequency plans according to the predicted hourly behaviour of the *MUF*.

Conversely, the *LUF*, apart from the geometry of the radio link and ionospheric conditions, is also dependent on the hardware characteristics and on the performance of the radio system. Therefore the hourly behaviour of *LUF* values can be decreased by changing the power supplied to the transmitter, the antenna gain, and according to the quality of the required service as summarized in the terms of Eqs. 9.2.36 and 9.2.37

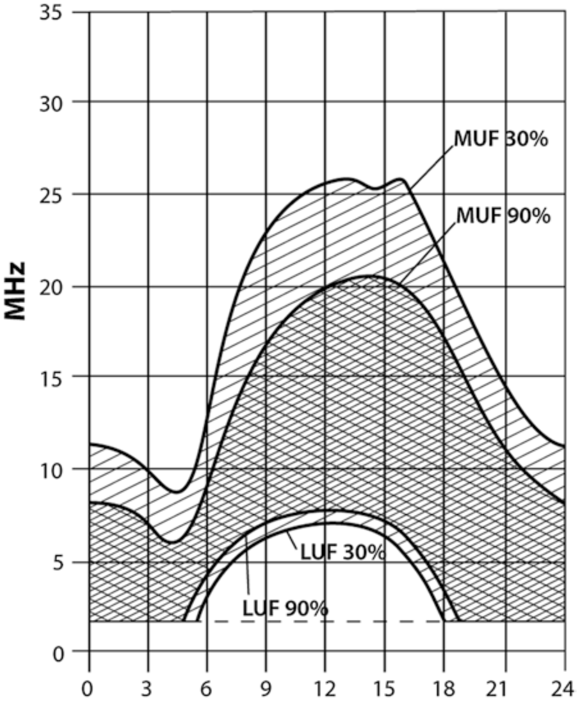


**Fig. 9.13** Example of the hourly monthly median prediction for a given distance and a given month provided by the French CNET (Committee National des Etudes Telecommunications) for two different values of probability: *MUF* (*upper graphs*) and *LUF* (*lower graphs*)

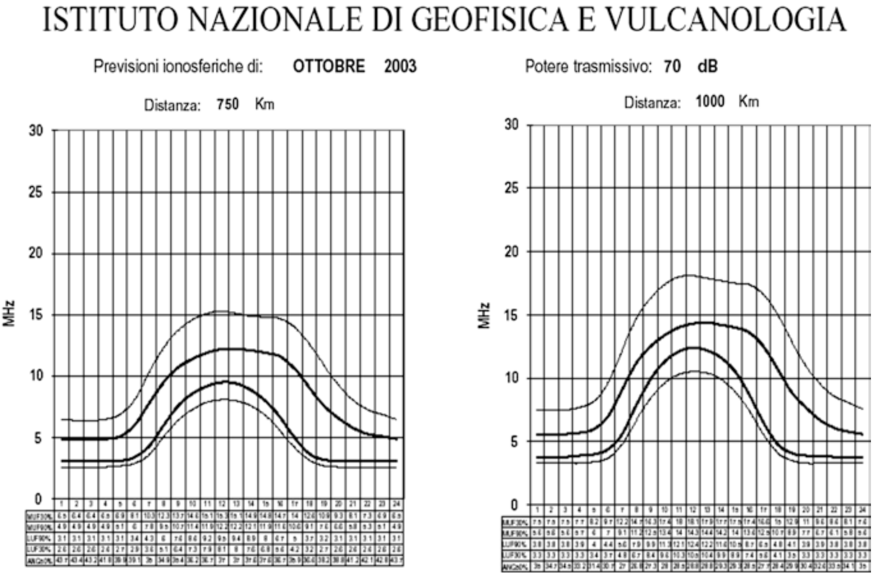


**Fig. 9.14** Example of the hourly median prediction of *MUF* and *LUF* for October 2003, with two values of probability, 90 % (in *bold*) and 30 %, and for two different distances, 750 km and 1,000 km

Figure 9.14 shows an example of monthly predictions provided by the Ionospheric Prediction Service of the Istituto Nazionale di Geofisica e Vucanologia (INGV), based on the same method developed by the French CNET, for a pair of standard distances, 750 km and 1,000 km, a fixed value of $P_T$, and valid for the Mediterranean region. If the distance is increased, the *MUF* and *LUF* values also increase while the angles of radiation decrease.

The ionospheric prediction services provide not only long-term predictions for a generic month and a standard distance valid for a region, but also monthly and short-term predictions for given point-to-point communications around the globe, for a given epoch, year, month, day, and radio equipment. See the examples of the Australian Ionospheric Prediction Service in http://www.ips.gov.au. In the past the main reference for HF point-to-point frequency planning was the monthly prediction, due to the obvious need to establish radio frequencies valid for at least one month. More recently, to satisfy the operative needs of civilian and military users and thanks to the opportunities provided by Internet communication, the main references are short-term predictions or forecasts, and real-time updates or nowcasts of ionospheric conditions together with working frequencies. As described in Chaps. 6 and 8, the updating of the climatological background, expressed as monthly median conditions, for short-term forecasting or real-time nowcasting, is achieved by complex procedures that use as input data both real-time measurements and ionospheric parameters with effective solar indices.

Often output is published as an alert with a percentage evaluation of an increase or decrease in hourly *MUF* values for given coordinates of a radio link. In the European

Upper Atmosphere Digital Service (DIAS service, http://www.iono.noa.gr/Dias) the hourly nowcast of *MUF* is issued for a given hour based on the nowcast map. In addition to the DIAS Service there is a publication that can be considered a real-time picture of the ionosphere over a region where a number of vertical ionospheric soundings are operating. The 15 min real-time f plot of the *foF*2 in comparison with the monthly median behaviour modelled by SIRM above a vertical sounding point provides real-time ionospheric conditions (Fig. 9.15).

If the measurements are very close to the monthly median prediction, this is proof of good performance by the mapping and long-term prediction procedure. The yellow line in Fig. 9.15 $f_{min}$ is the minimum frequency recorded in the ionogram and roughly represents the level of radio noise and absorption that occurs in the lower region of the ionosphere. This nowcast can be extremely useful for a particular type of point-to-point radio link, the so-called near vertical communication. The Near Vertical Incidence Sounding (NVIS) case is simply HF sky wave communication

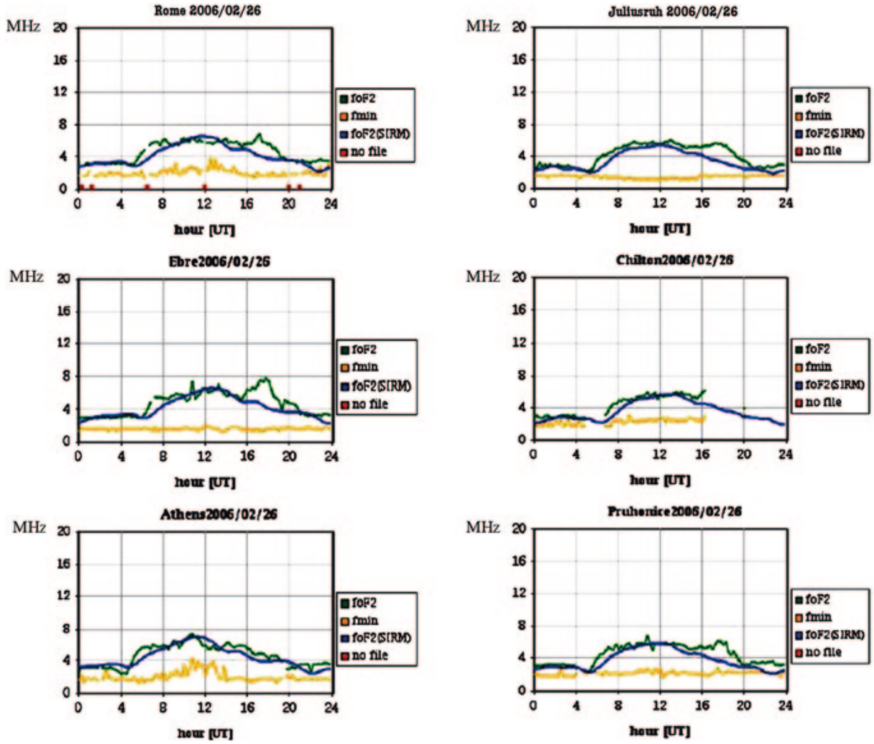


**Fig. 9.15** The f plots at different ionospheric stations in Europe provided by the DIAS service. The *blue lines* represent the predicted monthly median behaviours of the hourly values of the critical frequency *foF*2 by SIRM, while the *green lines* represent the real-time measurements obtained by the ARTIST autoscaling procedure of the same characteristics

at very short ranges, approximately the same as the ionosonde results for vertical incidence reflection. This kind of radio link is useful when it is necessary to connect two nearby points, a few dozen kilometres apart but separated by mountains or just with rough ground morphology, a typical application for military and civil defence.

### 9.2.6 HF Area Prediction and Nowcasting

The skip distance is defined as the minimum distance *D* for which it is possible to establish a radio link for a given frequency $f_{ob}$ and the critical angle of incidence $\varphi_0$ given by Eq. 9.2.1 (Fig. 9.3). For this distance *D* and this angle of incidence $\varphi_0$ the relative frequency *f* is also the maximum frequency reflected from the ionosphere according to Eq. 9.2.6. If a line is traced around a point of transmission at the same distance as the skip distance, that line is the isoline of the maximum frequency reflected by the ionosphere, or in other words the isoline of a given *MUF*. In Fig. 9.16, for a point of transmission located in Barcelona the isoline of the frequency 4 MHz is the skip distance and *MUF* for those points. The reflections occur at the mid point between the line and the transmitter. Inside the 4 MHz line the angles of transmission are so acute that reflections for hours 03, 04 and 05 UT are impossible. Consequently, within this area this frequency is not received at any point by sky waves but only very locally by ground waves. The 4 MHz frequency

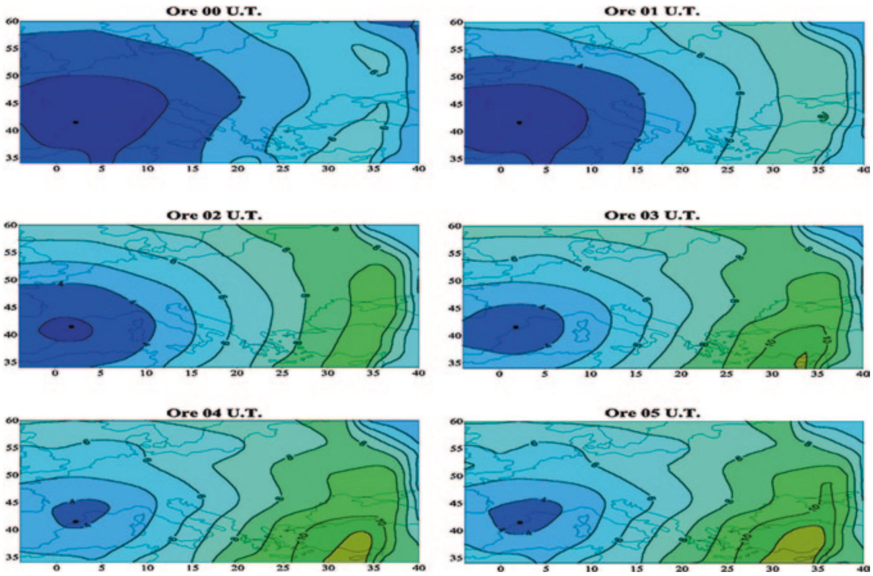


**Fig. 9.16** Example of isolines of the *MUF* in MHz or the skip distance variable with time for a point of transmission located in Barcelona. This sequence of maps, from 00UT to 05UT, clearly shows the effect of the Sun increasing the electron density from the South East

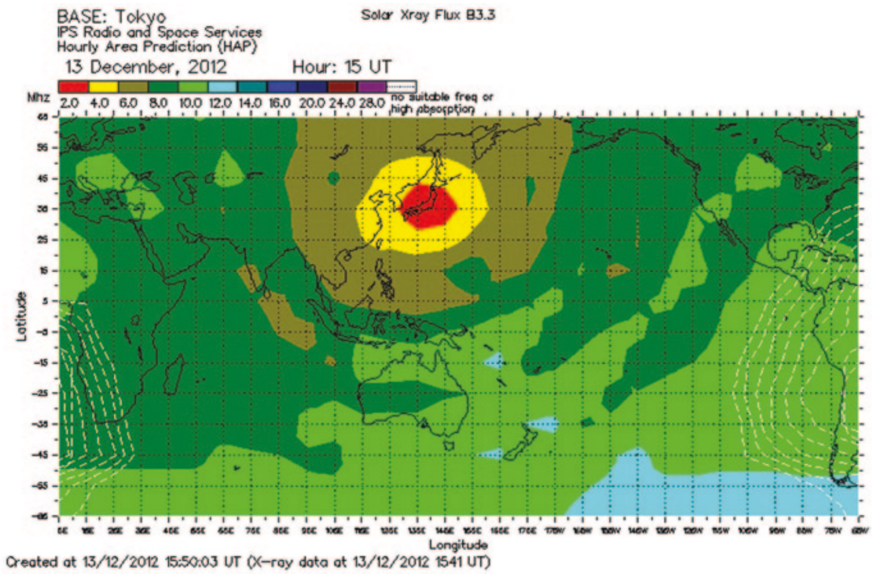


**Fig. 9.17** Nowcast optimum recommended frequency for a transmitter in Tokyo provided by the Australian IPS worldwide service

can be clearly received beyond that isoline up to the limit that the characteristics of the radio equipment permit, these being transmitting power, absorption, etc., or in other words the *LUF*. This diagram is typical for radio broadcasting applications or for fixed point to mobile point radio links. It is interesting to note how the area covered by a given frequency varies radically according to the time of day.

In principle the curves of equal *MUF* or equal skip distances are obtained in the same way as for point-to-point radio links: starting from the prediction of global or regional mapping of *foF*2 and *M*(3000)*F*2, the *MUF* is calculated for each point and the distance from the transmitter for all the area considered. Then the isolines are drawn using contour software code. Finally, based on the climatological background a nowcast is generated including correction with an effective solar or ionospheric index according to the method used. Figure 9.17 shows a nowcast optimum recommended frequency map provided by the IPS for the entire globe.

## 9.3 Existing Prediction and Nowcasting Propagation Procedures

In the historical introduction it was mentioned how after the Second World War the national ionospheric prediction services provided many different methods to predict and plan HF ionospheric radio propagation for long distance radio links. Since the 1980s computer programs and personal computers enabled users to predict the *MUF* between two points of the planet and many other radio parameters, like *LUF*, radio noise, antenna gain, displaying the results in various ways.

Most of these methods are based on the same philosophy described above, with the same global mapping method provided by W.B. Jones and R.M. Gallet, and a solar index input to predict the future monthly behaviour, but they differ for performance improvements or the areas considered. Furthermore, these methods were developed for military requirements by military institutions, and so detailed descriptions of the methods are often only published in internal reports and are not always widely available. Nowadays there are numerous methods running on-line on Internet. It is worth mentioning some well known ionospheric prediction methods that can be used on-line, originating from the USA, Australia, and Europe.

### 9.3.1 IONCAP, VOACAP, and ICECAP

Traditionally the American method IONCAP (Ionospheric Communications Analysis and Prediction Program) was one of the most widely used methods around the world, as well as its more recent enhancements in VOACP and ICECAP, used to calculate HF propagation parameters at any location. For a detailed review of its evolution see C. Rush in 1986. *MUF*, *LUF*, and field strength are the main parameters calculated by IONCAP. The required inputs for this method are the coordinates of the transmitter and receiver locations, transmitter power, universal time, month, and sunspot number. The software code in FORTRAN is available from the National Geophysical Data Center (ftp.ngdc.noaa.gov/STP/IONOSPHERE/MODELS/IONCAP/).

Based on the same principles as IONCAP, the more recent VOACAP and ICEPAC (Ionospheric Communications Enhanced Profile Analysis & Circuit) methods perform better and are modified to take into account the broadcasting needs firstly of Voice of America and secondly of high-latitude northern hemisphere propagation. They suffer from the same limitations as monthly median statistics. Figure 9.18 shows prediction graphs from the ICEPAC method, which are also available on the web together with an excellent technical manual.

### 9.3.2 The IPS Advanced Stand Alone Prediction System

The Internet site of the Australian Ionospheric Prediction Service, a government institution, is an important source of information not only for radio propagation but also as regards the solar-terrestrial environment and space weather phenomena: www.ips.gov.au. The IPS method is known as the Advanced Stand

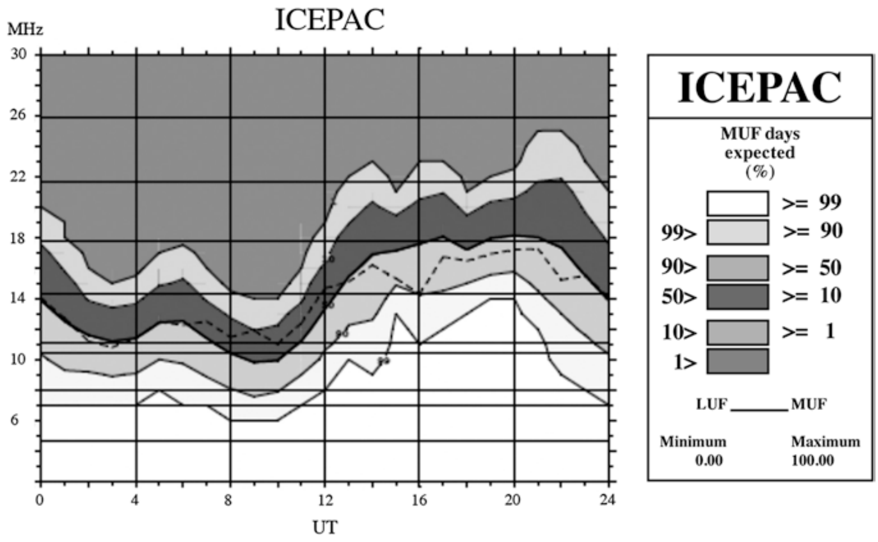
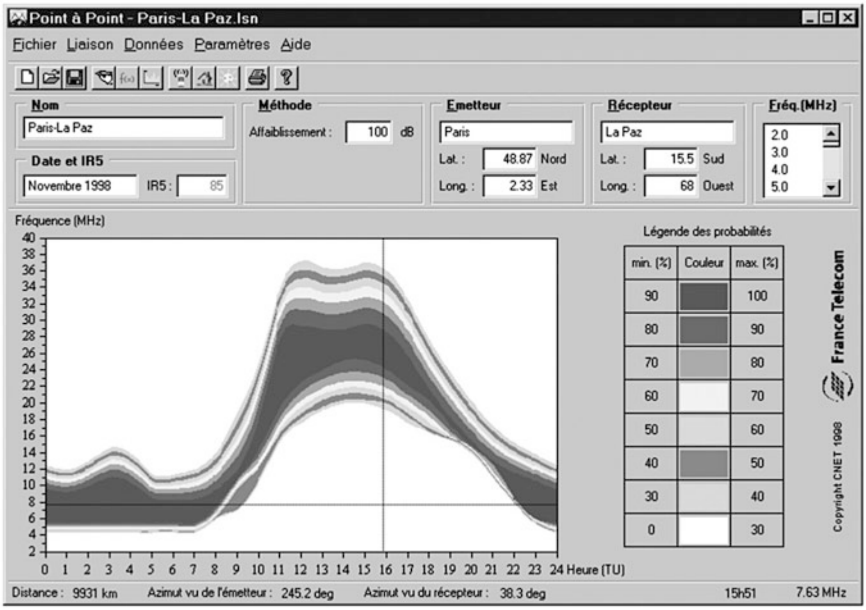


**Fig. 9.18** Long-term prediction of *MUF* for a long distance radio link: Rome to Los Angeles, calculated with the ICEPAC procedure for different values of probability (ftp.ngdc.noaa.gov/STP/IONOSPHERE/MODELS/IONCAP/)

Alone Prediction System (ASAPS) and is able to predict sky wave communication conditions for both the HF and VHF radio spectra. The method, developed by the IPS Radio and Space Services of the Australian Bureau of Meteorology (see Sect. 6.2.1.3), merges the features of the original IPS method with the ITU-R/CCIR models. Numerous graphic representations are based on this method according to the different needs of users. The graphs have regional or even global validity. Nowcasting maps also issue comparisons of median propagation conditions in terms of the ionospheric/solar index *T*.

## *9.3.3 France Telecom Method*

In Europe different methods were developed during the 1950s and 1960s. In Germany the FTZ method was specially designed for broadcasting purposes and adapted for use in conjunction with IONCAP as described in CCIR Report 894 in 1982. In France, first the SPIM (Service des Prevision Ionospheric Militaire), and then France Telecom developed methods. The French method was designed to provide median monthly predictions over a given region with different values of probability as shown in Fig. 9.19. The predictions are based on the binomial variation of the median monthly values of *foF*2, measured at a typical ionospheric station in the region considered, as a function of the solar sunspot index *IR*5. This index is similar to $R_{12}$ and provides a running mean of the monthly sunspot number over



**Nom**
Paris-La Paz

**Date et IR5**
Novembre 1998 IR5 : 85

**Méthode**
Affaiblissement : 100 dB

**Emetteur**
Paris
Lat. : 48.87 Nord
Long. : 2.33 Est

**Récepteur**
La Paz
Lat. : 15.5 Sud
Long. : 68 Ouest

**Fréq.(MHz)**
2.0
3.0
4.0
5.0

Légende des probabilités

| min. (%) | Couleur | max. (%) |
|---|---|---|
| 90 | ■ | 100 |
| 80 | ■ | 90 |
| 70 | ▒ | 80 |
| 60 | □ | 70 |
| 50 | ░ | 60 |
| 40 | ▒ | 50 |
| 30 | ░ | 40 |
| 0 | □ | 30 |

Distance : 9931 km Azimut vu de l'émetteur : 245.2 deg Azimut vu du récepteur : 38.3 deg 15h51 7.63 MHz

**Fig. 9.19** Example of point-to-point monthly prediction of *MUF* and *LUF* hourly conditions, given for 8 intervals of probability provided on-line by France Telecom at http://www.iono.enst-bretagne.fr. (Courtesy France Telecom)

5 months. However, long distance HF radio prediction graphs are also provided on-line by France Telecom as shown in Fig. 9.19 (www.iono.enst-bretagne.fr).

## 9.4 What Purpose Do Ionospheric Prediction, Nowcasting, and Forecasting Serve?

As shown in the previous chapters, the characteristics of an ionospheric propagation channel, whether it is HF or trans-ionospheric, are highly variable on timescales ranging from the 11 years of a solar cycle, to a few seconds. Even during its quietest periods, the Sun produces electromagnetic radiation and the solar wind, which simultaneously affect a variety of geomagnetic and ionospheric phenomena (Chaps. 2, 3 and 4), producing hour-to-hour changes in channel characteristics. Long-term variations, such as those induced by the solar cycle, are effectively described in this book by the long-term prediction programs currently in use (Chaps. 6 and 9). Very rapid changes require a high level of attention and skill by radio equipment designers to deal with them. In principal these can be handled by well trained radio operators during their management of channel frequencies. However, such capable operators are becoming increasingly rare. Consequently, users of HF and trans-ionospheric radio systems need assistance with their seasonal frequency planning and day-to-day frequency management (Chaps. 8 and 9).

Some users apply real-time channel evaluation (RTCE) techniques in their system management; even though it remains clear that there is no short-term ionospheric forecasting that can compete with advanced RTCE techniques for a fully adaptive radio system. Ideal usage of the ionospheric propagation channel cannot be achieved using RTCE alone because it provides information on existing conditions but not those likely to develop over the next few hours or days (see Chap. 8). Moreover, real-time usage of RTCE gives information on the locations of the transmitting and receiving equipment, which in the case of military and security applications is never completely acceptable. Users without ionospheric prediction and forecasting information may be badly affected when dramatic $foF2$ and $MUF$ depressions, polar cap absorption (PCA), or shortwave fadeouts (SWF) cause a total or partial loss of communication. As a result, ionospheric forecasting, nowcasting specification, and long-term predictions play an important role in frequency planning and management procedures.

## Suggested Readings

Bradley PA (1989) Propagation of radio waves in the ionosphere. Radio wave Propag IEEE Electromagn Ser 30:95–113

Breit G, Tuve MA (1926) A test of the existence of the conducting layer. Phys Rev 28:554–575

Budden KG (1961) Radio waves in the ionosphere. Cambridge University Press, Cambridge

CCIR Report (1982) CCIR interim method for estimating sky-wave field strength and transmission loss at frequencies between the approximate limits of 2 and 30 MHz. Propagation in Ionized Media, Recommendations and Reports of the CCIR, Rep. 252-2, VI, International Telecommunication Union Geneva

Davies K (1990) Ionospheric Radio. IEE Electromag Waves Ser 31, Peter Peregrinus Ltd, London

Galkin IA, Reinisch BW (2008) The new ARTIST 5 for all Digisondes. Iono Net Ad Group Bull 69:1–8

Halley P (1970) Propagation des ondes radioelectriques dans l'enviroment terresrre. Internal report of the Centre National d'Etudes des Telecommunications; Division des Previsions Ionospheriques, 2$^{eme}$ Fascicule

McNamara LF (1991) The Ionosphere: Communications, Surveillance, and Direction Finding. Krieger Publishing Company, Malabar

Martyn DF (1935) The propagation of medium radio waves in the ionosphere. Proc Phys Soc 47:323–339

Ochs A (1970) The forecasting system of the Fernmeldetechnischen Zentralamt (FTZ). Ionospheric Forecasting, NATO-AGARD Conference Proceedings 49. paper 43

Radio Science Bulletin no. 327 Special sections honoring Jenifer Haselgrove, 2008

Rawer K (1975) The historical development of forecasting methods for ionospheric propagation of HF waves. Radio Sci 10:669–679

Rush CM (1986) Ionospheric radio propagation models and predictions—a mini-review. IEEE Trans Antennas Propag AP-34:1163–1170

Sizun H (2005) Radio Wave Propagation for Telecommunication Applications. Springer, Heidelberg

Smith N (1939) The relation of radio sky-wave transmission ionosphere measurements. Proceedings of IRE 27:332–347

# Additional Web Sites

http://www.iono.noa.gr/Dias
http://www.iono.enst-bretagne.fr
http://www.ips.gov.au
ftp.ngdc.noaa.gov/STP/IONOSPHERE/MODELS/IONCAP/
http://www.ips.gov.au/IPSHosted/INAG/web-69/2008/artist5-inag.pdf

# Chapter 10
# Current and Future Trends in Ionospheric Prediction and Forecasting

## 10.1 Introduction

One of the recent geophysical events that attracted considerable scientific and public attention was the geomagnetic storm of March 1989. This storm occurred near the solar maximum of the 22nd solar cycle, with a sunspot number $R_{12}$ of about 160. It was marked by an enhancement of the ring current and quantified on the *Dst* index with minimum hourly values of $-599$ nT on 13 March 1989. The storm effect was preceded by a Sudden Impulse (SI) of the A type at 0127 UT. The initial phase of the geomagnetic disturbance, about 1 h in duration, was followed by a long–lasting main phase reaching its maximum at approximately 2100 UT on the first day of the storm, qualifying this event as extreme.

The overall impact during this event was summarized in literature as follows: (1) LORAN maritime navigation systems outages on 13 March 1989; (2) communications circuit anomaly on GOES-7 on 12 March 1989; (3) loss of imagery and communications on GOES-7 on 13 March 1989; (4) three low-altitude NOAA polar orbiters and a DMSP polar orbiter had trouble unloading torque; (5) telecommunications satellite CS-3B failed on 17 March 1989; (6) MARECS-1 satellite experienced operational problems; (7) series of seven commercial geostationary satellites had many problems maintaining operational attitude; and most importantly (8) a transformer failure on one of the main power transmission lines in the Hydro-Quebec system precipitated a catastrophic collapse of the entire power grid when 6 million people lost electrical power for 9 or more hours.

Since the March 1989 event our society has continued to progress and develop an ever more sophisticated infrastructure, and its dependence on the properties and behaviour of the solar-terrestrial system has become increasingly sensitive. The reliance on space-based technologies for communication, navigation, and resource management is growing rapidly and with it our vulnerability to disruptions caused by extreme geophysical events along with secular atmospheric evolution. The future commercialization of space, with business enterprises to be set up in low Earth orbit within the thermosphere and ionosphere, will be a societal issue

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1_10,

in years to come. Effects of extreme solar activity on other planets, and elsewhere in the solar system, can also be important, and violent flare activity can even be observed from other stars.

This portion of the extended geospace has played a vital role in maintaining the habitability of our home planet and human well-being in general, and is the focus of most financial investment in operational space-based activities requiring: (1) knowledge of the Sun and solar-terrestrial interactions, as well as the physics and chemistry of the ionosphere (Chap. 2); (2) ionospheric measurement techniques and the role of major ionospheric stations in this context (Chap. 3); (3) modelling capabilities of ionospheric processes (Chaps. 4, 5, and 7); (4) mapping methods for the main ionospheric characteristics (Chaps. 6 and 7); (4) prediction and forecasting supported by international collaboration within ionospheric COST actions, IRI activities, and different EU and ESA research projects (Chaps. 8 and 9).

Consequently, current and future trends in ionospheric prediction and forecasting must contribute to an improved understanding and effective mitigation of the impact of significant geophysical events on our technologically reliant society, through the development of observational networks and complete ionospheric-thermospheric models. In this respect, Sect. 10.2 introduces ionospheric prediction and forecasting in radio wave propagation as an important contribution to space weather. Section 10.3 deals with specific issues such as mitigation of disturbances and signal errors in GNSS communication and surveillance systems. Lithosphere and ionosphere coupling, considered in Sect. 10.4, is an object of increasing interest among the scientific community as a consequence of the recent dramatic events. The concluding remarks of this book are set out in Sect. 10.5.

## 10.2 Ionospheric Prediction and Forecasting in Radio Wave Propagation as an Important Contribution to Space Weather

In antiquity the term "meteorology", from the ancient Greek *μετεωρολογικά* or from the ancient Roman *meteorologica*, was first used by Aristoteles (340 B.C.) to describe all the phenomena that occur between the Earth and the sky. In the modern epoch the meaning of Meteorology has generally been considered as the science that explains, describes, and predicts the phenomena occurring mainly in the troposphere. Within this thin but dense gaseous layer of about 10 km average altitude, ranging from 8 km over the poles and about 12 km over the equator, the interaction between the components of the atmosphere is more thermodynamic than electromagnetic.

During the last decade of the 20th and the first decade of the current 21st century the observation and study of phenomena in space and specifically in the near-Earth environment have greatly increased in importance due to the mass adoption of technological applications. According to the McGraw-Hill Science and Technology Dictionary, one of the definitions of space weather provided by the US National Science Foundation is: “Conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health”.

This definition embraces a large number of subjects ranging from the nuclear physics of the Sun, to the technological aspects of spacecraft and ground-based systems, to the biological effects on living beings. The second  Chapter of this book briefly described the wide variety of geophysical phenomena related to the interaction between the Sun and the space around the Earth, defined as solar-terrestrial interactions, and the subsequent chapters considered the Earth’s ionosphere and ionospheric radio propagation.

Prediction and forecasting of ionospheric phenomena and applications for technological systems like telecommunications or navigation was certainly part of space weather before this term became established.

Prediction needs to take into account both the possibility of extreme events, as occurred in the last two centuries in 1859 and in 1921, before and at the beginning of the radio era, and more recently in March 1989 and October 2003, as well as long-term or very slowly changing phenomena. For example the cooling and lowering of the ionospheric regions due to a possible inverse greenhouse effect and an additional proof of global warming, or simply due to the influence of the variability of the geomagnetic field on ionospheric characteristics and parameters.

Long-term alterations in the maximum electron density height caused by global changes, as well as short-term changes due to solar radiation and geomagnetic storms, can determine an expansion or lowering of the ionosphere with potentially dramatic effects on low orbit (LEO) satellites.

Solar radiation storms, geomagnetic storms, and radio blackouts are the three categories of phenomena considered by NOAA on a numerical scale, similarly to hurricanes or earthquakes, giving a rough quantitative estimate of their effects. Table 10.1 shows the scale of radio blackouts caused by the effects of space weather on ionospheric dynamics.

## 10.3 Mitigation of Disturbances and Signal Errors in GNSS and Other Systems

Technological systems particularly sensitive to the geophysical events discussed in this book include the following.

### *10.3.1 Navigation Systems*

As the accuracy of long-range (LORAN)/DECCA and OMEGA maritime navigation systems depends on precise knowledge of the altitude of the bottom of the ionosphere, any vertical changes in the ionospheric D region during solar

**Table 10.1** Radio blackout scale (from NOAA http://www.swpc.noaa.gov/NOAAscales/)

| Radio black outs scale | HF radio | Navigation | Number of events |
|---|---|---|---|
| Minor | Weak or minor degradation of HF radio communication on sunlit side, occasional loss of radio contact | Low-frequency navigation signals degraded for brief intervals | 2,000 per cycle (950 days per cycle) |
| Moderate | Limited blackout of HF radio communication on sunlit side, loss of radio contact for tens of minutes | Degradation of low-frequency navigation signals for tens of minutes | 350 per cycle (300 days per cycle) |
| Strong | Wide area blackout of HF radio communication, loss of radio contact for about an hour on sunlit side of Earth | Low-frequency navigation signals degraded for about an hour | 175 per cycle (140 days per cycle) |
| Severe | HF radio communication blackout on most of the sunlit side of Earth for on to two hours. HF radio contact lost during this time | Outages of low-frequency navigation signals cause increased error in positioning for one to two hours. Minor disruptions of satellite navigation possible on the sunlit side of Earth | 8 per cycle (8 days per cycle) |
| Extreme | Complete HF radio blackout on the entire sunlit side of the Earth lasting for a number of hours. This result in no HF radio contact with mariners and en route aviators in this sector | Low-frequency navigation signals used by maritime and general aviation systems experience outages on the sunlit side of the Earth for many hours, causing loss in positioning. Increased satellite navigation errors in positioning for several hours on the sunlit side of Earth, which may spread into the night side | Less than 1 per cycle |

flare effects and/or geomagnetic storms can introduce errors of several tens of kilometres in location determination. Hence the SID monitor network, a low-cost set of sensors for lower ionosphere layers, was recently presented to the scientific community by the Stanford University group, offering good prospects for continuous monitoring of ionospheric activity (http://sidstation.loudet.org/stations-list-en.xhtml).

Significant positioning errors result when signals are changed by sudden total electron content variations and ionospheric irregularities of different kinds, causing a loss of phase lock, data loss, and consequent loss of position update. An

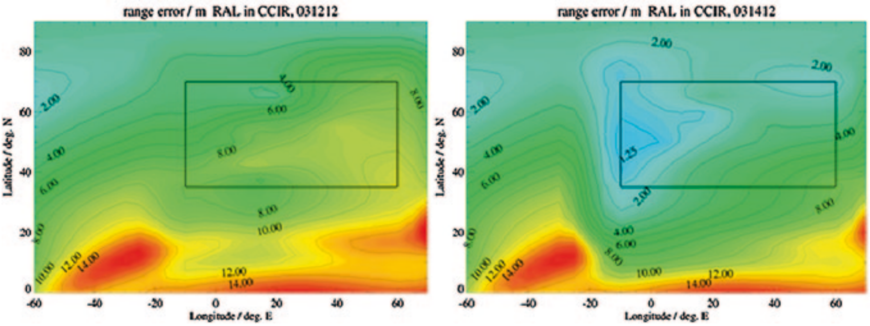


**Fig. 10.1** Range errors over the European area on the storm day of 12 March (*left*) and a quiet day on 14 March (*right*) 1989 at 1200 UT (Courtesy R. Leitinger 2002)

example of range error as the first order vertical ionospheric plasma influence on the main GPS signal ($f = 1575.42$ MHz) is given in Fig. 10.1 for the March 1989 storm. This is the error "seen" by a single frequency user for overhead GPS satellites. Slant range error can be higher by a factor of up to 5, depending on zenith angle and *TEC* gradients, while without gradients the factor is up to 3.

The range errors within the marked rectangle are the errors calculated with the RAL short-term ionospheric forecasting maps of *foF2* and *M*(*3000*)*F2* (see Chap. 8). The rectangle is surrounded by a buffer zone. Outside the buffer zone the data were calculated from the ITU-R maps for *foF2* and *M*(*3000*)*F2* and therefore there is no difference between the two days. The differential range errors calculated between 2 days, one storm day on 12 March and one quiet day on 14 March 1989, are shown in Fig. 10.2. These significant differences are a major concern for satellite navigation systems.

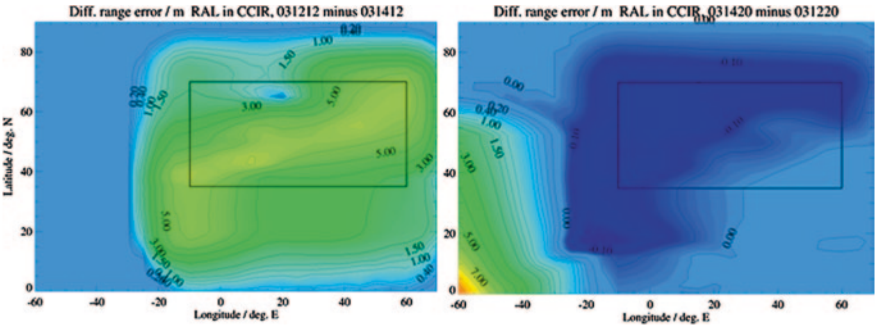


**Fig. 10.2** The differential range errors over the European area on the storm day of 12 March and a quiet day on 14 March 1989 at 1200 UT (*left*) and 2000 UT (*right*) (Courtesy R. Leitinger 2002)

Since satellite navigation now encompasses wide and growing applications it has become necessary to switch from military controlled systems (GPS and GLONASS) to a GNSS system under complete civil control. The first step was the implementation of GNSS-1 using Geo stationary satellites to improve the positioning accuracy available from GPS and GLONASS alone. The European contribution to GNSS-1 is European Geo stationary Navigation Overlay System (EGNOS). The development of the operational EGNOS system by the ESA involves the EGNOS IONO modelling and performance/integrity aspects and the EGNOS IONO "worst case" test scenario against integrity and performance objectives. The EGNOS system with its Ranging and Integrity Monitoring Stations (RIMS) could also be used by space weather services as a data provider. After completion of GNSS-1, a GNSS-2 system is foreseen that will be under complete civil control.

### *10.3.2 Communication Systems*

Prediction and forecasting accuracy depend on the complexity of the ionospheric structure and the propagating frequency of the radio waves of: HF communications and broadcasting, UHF (ultra high frequency 300–3000 MHz)/SHF (super high frequency 3–30 GHz) Satcoms, LEO and MEO cellular and data SATCOM, VLF-MF communications and broadcasting. Ionospheric irregularities, multipath and attenuation of signals as well as the Doppler effect can cause a complete blackout at HF, HF Doppler shifts of up to 30 Hz, HF multipath up to 8 ms, 30 dB fades at UHF and 20 dB fades at L-band frequencies. Consequently, data/symbol errors and loss of radio wave/ground wave interference communications are always possible. However, by using ionospheric prediction and forecasting in conjunction with the technical elements and characteristics of radio systems, the prediction of performance of ionospheric-dependent radio systems can be successfully determined as described in Chap. 9.

To demonstrate this connection in the case of an HF system, the critical frequency *foF2*, corresponding to the peak electron density in the F region *NmF2*, is presented in Fig. 10.3 one day before, during, and two days after the storm (12–16 March 1989) at the Slough ionosonde station. The March 1989 storm generally produced similar *foF2* variations at all examined mid-latitude ionosonde stations and provided an excellent example of long-term negative storm effects. Furthermore, on the day before and shortly after the storm, related effects are seen, with the hourly *foF2* values closely following the monthly median values. This type of event can be successfully predicted because the morphology of storm effects, under such a clearly negative storm pattern, was well described by Prölss (1995).

Oblique incidence sounding data (see Sect. 3.3.2) play an important role in controlling the real characteristics of HF propagating signals in the ionospheric radio channel, and for developing a capability to provide reliable HF frequency management and predictive services for essential circuits. Accordingly, experimental studies with oblique incidence ionosondes were conducted over a number of paths

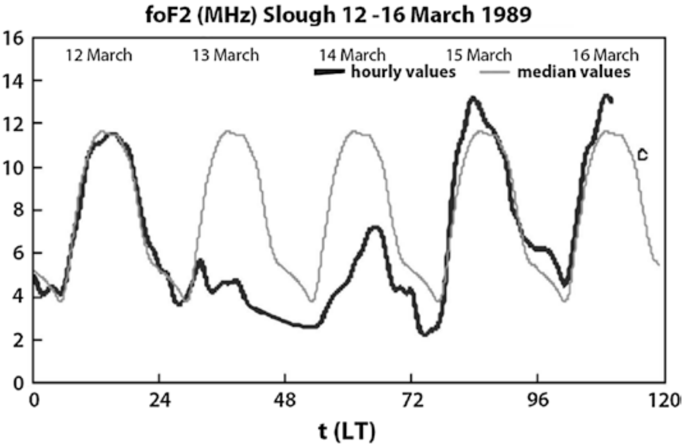


**Fig. 10.3** The *foF2* variations at the Slough (51.5°N, 359.4°E) ionosonde station during the March 1989 storm

together with vertical sounding campaigns during the March 1989 event. From a typical oblique sounding ionogram it is possible to determine the Maximum Observable Frequency (*MOF*), which is the highest frequency at which signals transmitted from a sounder are observed on an ionogram regardless of the propagation path involved, and the Lowest Observable Frequency (*LOF*), which is the lowest frequency at which signals transmitted from a sounder and propagated via the ionosphere are observed on an ionogram.

The *LOF* and *MOF* variations in the ionospheric path Kiruna (67.8° N, 20.4° E)—Lycksele (64.6° N, 18.8° E) during 12–15 March 1989 are shown in Fig. 10.4. It can be seen that there is no reflection during the hours of the main

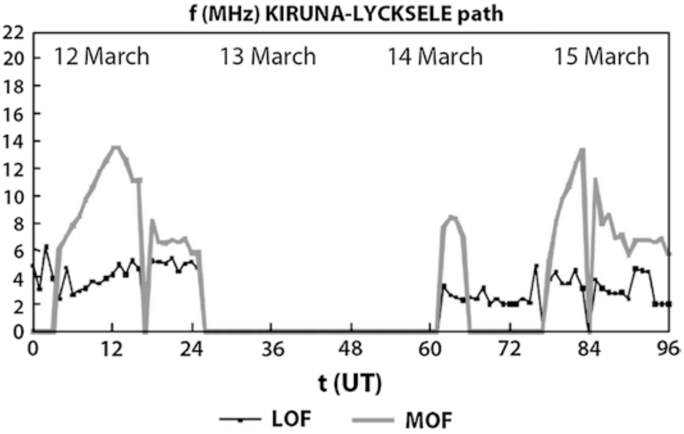


**Fig. 10.4** The *LOF* and *MOF* variations during 12–15 March 1989

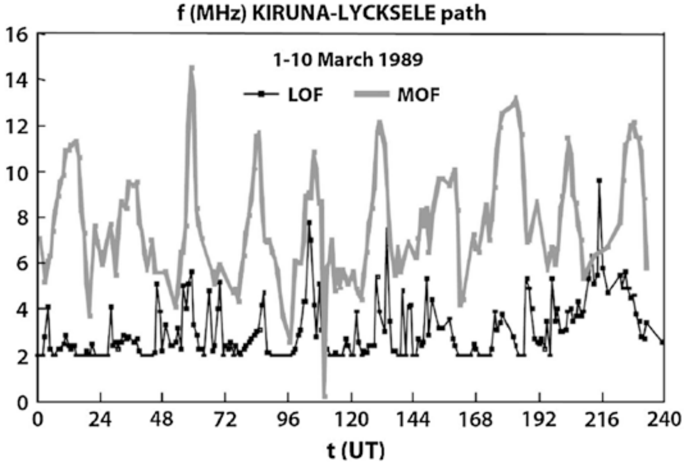


**Fig. 10.5** The *LOF* and *MOF* variations during 1–10 March 1989

phase of the geomagnetic storm, neither in the lower nor higher parts of the ionosphere. The regular ionospheric structure completely disappeared and hence the dramatic consequences for propagation conditions.

The *LOF* and *MOF* variations on the same ionospheric path Kiruna (67.8° N, 20.4° E)—Lycksele (64.6° N, 18.8° E) during the first 10 days of March 1989 are shown in Fig. 10.5. These variations are a good representation of ionospheric climatology observed from the oblique sounding results, before the storm event.

**Fig. 10.5** The *LOF* and *MOF* variations during 1–10 March 1989

The data obtained from the Swedish networks used here reveal the features of frequency parameters such as *LOF* and *MOF* during ionospheric disturbances. From the sequences of range-frequency characteristics related to the single intense storm and to quiescent conditions, it is seen that increased short-term activity can be interpreted as a precursor for more disturbed periods that may follow.

### 10.3.3 Surveillance

Total electron content variations influence UHF/SHF radars, HF OTHR (Over the Horizon Radar), space-based SAR (Syntetic Aperture Radar) and geo-location. The range and bearing errors, spectral distortion, loss of target discrimination, and loss of phase coherence across SAR aperture can completely prohibit remote sensing and/or involve many km of range uncertainty at HF, over 200 m at UHF, and HF bearing errors up to 100°.

For satellite radar altimeters intended to perform space measurements of the distance between the satellite and the sea surface height using 13 GHz frequency radio waves, the effect of ionospheric modification is typically in the order of 30 cm. This magnitude is comparable to the amplitude of large oceanic signals.

In order to achieve the goal of an *RMS* less than 2 cm in altimetry, it is essential to use an adequate empirical, statistical, and physical model, or a mix of all the plasmasphere-ionosphere-thermosphere models, or use the Doris or GPS dual frequency measurements to estimate highly accurate *TEC* maps.

## 10.4 Lithosphere and Ionosphere Coupling

The study of the geophysical phenomena that appear prior, during, and after a seismic event has increased considerably over recent decades. This has been the object of research of numerous scientists from different disciplines, including, in addition obviously to seismology, also hydrology, geomagnetism, atmospheric physics, geochemistry, as well as ionospheric radio propagation. Experimental observations both from ground-based measurements and from space satellites have been performed or are in progress to study coupling of the lithosphere and ionosphere, with the possibility of determining an ionospheric precursor to earthquakes.

These studies have produced a large number of papers published in scientific literature and presented in special sessions at international conferences both for seismologists and ionospheric scientists, including the URSI and the European Geophysical Union General Assemblies.

For a complete treatment of this argument and a long list of references, see the volume published by Pulinets and Bordaciuk (2004), and more recently the review by Liperovsky et al. (2008). The Section presents a very short overview of this theme, delineating the main hypothesis for the mechanism that explains lithosphere-ionosphere coupling and the perplexities that animated the associated scientific debate.

Furthermore, the ionospheric effects of lithospheric phenomena could be considered together with a larger class of studies defined as electromagnetic effects of seismic events. For these studies it is important to mention the paper by Gokhberg et al. (1982), which reported wideband electromagnetic radiation occurring just prior to an earthquake of grade 7 on the Richter scale. After a period of relatively quiet levels of radio noise at 81 kHz, an anomalous amplitude increase of 15 dB above the normal level was recorded about one hour before the main shock in Japan on 31st March 1981. The radio noise amplitude returned to normal levels just after the event. Anomalous variations in VLF (3–30 kHz), and ELF (<3 kHz) emission associated with strong earthquakes above grade 6 on the Richter scale are reported by various authors.

As regards traditional ionospheric parameters, anomalous variations in ionospheric characteristics or total electron content, appearing simultaneously or after a seismic event have been noted since 1969. A disturbance in the ionosphere caused by the earthquake at Kurile Islands on 11th August 1969 was reported by Weaver et al. (1970). It is interesting to note their explanation that the vertical components of the longitudinal pressure waves produced by seismic waves moved the contours of the constant electron density layers. The AGWs (Atmospheric Gravity Waves) that propagate upward, exhibit increasing amplitude as the density of the neutral atmosphere decreases, so that displacements in the order of millimetres on the Earth's surface extend to displacements in the order of kilometres at ionospheric heights.

The mechanism of AGWs, beside the well known TIDs (see Sect. 2.5.1) excited by weather fronts in the lower atmosphere, are not only observed after an earthquake but are also considered as precursors of a seismic event at short and even long distances. The source of AGWs are probably long period oscillations of the ground, or the sea in the case of a tsunami, that propagate into the ionosphere, as emphasized by V. A. Liperovsky.

Over the last decade of the 1900s S. Pulinets and K. Bordachuk developed a physical model to explain some anomalous variations in the ionosphere, taking into account and correlating ionospheric measurements and gas emission over the ground in a seismogenetic area. The emanation from the ground of various chemical substances like radon, and light gases including helium and hydrogenous as well as submicron and metallic aerosols, changes the electrodynamic properties of the atmosphere over a region originating an earthquake. The radon released from the crust is an important source of ionization and clusters of heavy ions in the atmospheric layer close to the ground produce an additional electric field overlapping the total (ionosphere-ground) electric field.

The generation of an electric field due to the electrode effect in the atmosphere, at ground level leads to small scale irregularities in the E region and large scale irregularities in the upper part of the ionosphere. For details and the recent evolution of this model see the above cited volume.

Concerning future trends in these investigations, it is important to add that experimental projects based on space satellite measurements are in preparation, sponsored by international organizations. Worth mentioning is the DEMETER satellite experiment sponsored by the French CNRS and led by M. Parrot. This satellite is dedicated to studying seismo electromagnetic effects in the ionosphere and is equipped to detect ion composition, electron density and temperature, energetic particles, and radio emissions over a wide frequency range ELF-VLF, which are considered precursor phenomena occurring a few hours before seismic events and volcanic eruptions. Unlike ground-based ionospheric vertical soundings, which might be limited to one point, or be too far from the epicentre, satellite measurements can cover large seismic areas, although the measurements need to be validated against those taken on the Earth's surface. A Langmuir probe, electric and magnetic sensors, and an energetic spectrum analyzer are utilized (see the recent results on ionospheric density variations recorded before the 2010 magnitude 8.8 earthquake in Chile). Sea surface monitoring by the HF OTHR for an early tsunamis alert can be considered in this context.

It is now clear that electromagnetic seismic precursors and the possible effects on the ionosphere should not be neglected and deserve further investigation in the future in order to surpass their status among the numerous seismic precursors only recognized *a posteriori*. Numerous statistical analyses have been performed regarding variations from median hourly values classed as normal or quiet ionospheric conditions, raising the question as to whether the observed effects are strong enough to exceed normal ionospheric variability induced by other causes.

In other words, is it possible for similar effects to occur without any subsequent earthquake? For a concise critical opinion of this important and still debatable issue, see the paper published by Rishbeth (2006).

## 10.5 Conclusions

This book on Ionospheric Prediction and Forecasting is the result of three decades of fruitful and scientifically rewarding collaboration in research and joint projects by the authors. It presents a personal view of past, present, and future ionospheric prediction and forecast modelling and mapping from the perspective of their evolution during the last few decades in response to the needs of radio users implementing specific propagation applications. Therefore, the numerous references to relevant literature following each Chapter are of utmost importance for an in-depth knowledge of the subject. In that context it is important to note that symbols and letters corresponding to physical quantities have been maintained, where possible, the same as in the original references to facilitate comparisons and further considerations.

The material in this book demonstrates that in spite of continuous developments in modelling and increased quality and quantity of data, accurate ionospheric prediction and forecasting remains very difficult due to the extensive climatological ionospheric fluctuations that occur routinely. These have significant implications for radio propagation research, since precise forecasts of ionospheric conditions can be used to mitigate or avoid adverse effects on the communication and navigation systems that, notwithstanding diverse new technologies, still strongly depend on ionized media in the Earth's upper atmosphere.

Given that most scientific endeavours are never ending, with successful results leading naturally to new research themes to investigate, this literary journey can only be concluded by emphasizing that:

1. a major challenge for the twenty-first century, in order to achieve effective prediction and forecasting, is to define an integrated, multi-dimensional pattern of ionospheric changes, involving observation with standardized data formats and calibration procedures;
2. an advanced theory and modelling of the Earth's upper atmosphere is required, based on internationally coordinated, state-of-the-art real-time observational facilities and sophisticated computer simulation capabilities.

Currently, there are two distinct developments underway. The first focuses on efforts to refine and improve existing ionospheric models of all types beyond their climatological level, i.e. the average conditions observed during the forecast period. To date, a large quantity of ionospheric data, obtained from a diversity of ground- and space-based measurements, have been collected and analyzed with the aim of combining data and models into assimilative systems and/or constructing patterns of space weather events and effects that are most appropriate for a given location and specific user requirements, in order to predict or correct propagation effects in particular radio applications.

The second development is real-time use of data to provide more accurate specification of an instantaneous ionosphere. Improved techniques for solar-terrestrial monitoring and rapid advances in computer and digital data communications technology have created a number of possibilities for real-time data access and ionospheric database systems. In this respect it is absolutely vital for current and future developments that real-time data streams from DIAS, GIRO (Global Ionospheric Radio Observatory), and other global and/or regional networks provide at least hourly updated ionospheric data, together with a steady flow of relevant solar-terrestrial data worldwide.

## Suggested Readings

Axford WI, Hagfors TN (1996) 50 Years of ionospheric research at Max-Planck-Institut fur Aeronomie. In: Kohl H, Ruster R, Schlegel L (eds) Modern Ionospheric Science. European Geophysical Society, Katlenburg-Lindau

Bremer J (2004) Investigations of long term trends in the ionosphere with worldwide ionosonde observations. Adv. Radio Sci 2:253–258

Cander LjR, Mihajlovic SJ (1998) Forecasting ionospheric structure during the great geomagnetic storms. J Geophys Res 103(A1):391–398

Galkin IA, Reinisch BW, Huang X, Bilitza D (2012) Assimilation of GIRO data into a real-time IRI. Radio Sci 47:RS0L07. doi:10.1029/2011RS004952

Gokhberg MB, Morgounov VA, Yoshind T, Tomizawa L (1982) Experimental measurements of electromagnetic emissions possibly related to earthquakes in Japan. J Geophys Res 87(B9):7824–7828

Liu JY, Chen YI, Chuo YJ, Chen CS (2006) A statistical investigation of preearthquake ionospheric anomaly. J Geophys Res 111:A05304. doi:10.1029/2005JA011333

Liperovsky VA, Pokhotelov OA, Meister CV, Liperovskaya EV (2008) Physical models of coupling in the lithosphere-atmosphere-ionosphere system before earthquakes. Geomag Aero 48:795–806

Hapgood MA (2011)Towards a scientific understanding of the risk from extreme space weather. Adv Space Res 47:2059–2072

Jakowski N, Leitinger R, Ciraolo L (2004) Behaviour of large scale structures of the electron content as a key parameter for range errors in GNSS applications. Ann Geophys 47:1031–1047

Molchanov OA, Hayakawa M (1998) Sub ionospheric VLF signal perturbations possibly related to earthquakes. J Geophys Res 103:17,489–17,504

Parrot M (2006) First results of the DEMETER micro-satellite. Planet Space Sci 54(5):411–557

Píša D, Parrot M, Santolík O (2011) Ionospheric density variations recorded before the 2010 Mw 8.8 earthquake in Chile. J Geophys Res 116. doi:10.1029/2011JA016611

Prölss GW (1995) Ionospheric F-region storms. In: Volland H (ed.) Handbook of Atmospheric Electrodynamics. CRC Press, Boca Raton

Pulinets S, Boyarchuk K (2004) Ionospheric Precursors of Earthquakes. Springer, Heidelberg

Rishbeth H, Roble RG (1992) Cooling of the upper atmosphere by enhanced greenhouse gases: modeling of the ionospheric and thermospheric effects. Planet Space Sci 40:1011–1026

Rishbeth H (2006) Ionoquakes: earthquakes precursors in the ionosphere? Eos 87(32):8

Reinisch BW, Galkin IA (2011) Global ionospheric radio observatory (GIRO). Earth Planet Sci 63(4):377–381

Weaver PF, Yuen PC, Prölss GW (1970) Acoustic coupling into ionosphere from seismic waves of the earthquake at Kurile Islands on August 11–1969. Nature 226:1239–1241

Wright JM, Lewis FP, Corell RW, Thomas AR, Huntress WT, Devine JF, Waldon RE (1997) US National space weather program, The implementation plan. FCM-P31-1997 Washington, DC

# Additional Web Sites

SID Monitoring Station (2010) VLF Stations List. Available at: http://sidstation.loudet.org/stations-list-en.xhtml

SuperSID Manual (2009) Stanford Solar Center, Stanford University. Stanford, USA. Available at: http://solarcenter.stanford.edu/SID/Distribution/SuperSID/supersid_v1_1/Doc/SuperSIDManual_v1.pdf

# Appendix A

1. Notes on the vector *calculus*

A generic vector **a**, written in bold character, with $a_x$, $a_y$ and $a_z$ its components along the 3 Cartesian axes OX, OY, OZ, respectively, and **i**, **j** and **k** the unit vectors oriented like the same 3 axes is:

$$\mathbf{a} = a_x\,\mathbf{i} + a_y\,\mathbf{j} + a_z\,\mathbf{k} \tag{A.1}$$

The magnitude of vector **a** is:

$$|\mathbf{a}| = a = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{A.2}$$

The scalar or dot product between a vector **a** and another generic vector **b** with the components $b_x$, $b_y$ and $b_z$ is still a scalar quantity defined as:

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z = ab\cos\beta \tag{A.3}$$

where a is the magnitude of the vector **a**, b the magnitude of the vector **b** and β the angle between the directions of the two vectors.

The vector or cross product of two vectors **a** and **b** is a vector defined as:

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} = \left(a_y b_z - a_z b_y\right)\mathbf{i} + (a_z b_x - a_x b_z)\,\mathbf{j} + \left(a_x b_y - a_y b_x\right)\mathbf{k} \tag{A.4}$$

The magnitude is $c = ab\ \sin\ \beta$, where a and b are defined according to Eq. A.2, β is the smaller angle between the directions of the two vectors.

The equation:

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k} \tag{A.5}$$

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1,

defines the vector differential operator, $\nabla$, *del* or *nabla*.

The gradient of a scalar function S is $\nabla\mathbf{S}$ or grad S:

$$\nabla \mathrm{S} = \frac{\partial \mathrm{S}}{\partial x}\mathbf{i} + \frac{\partial \mathrm{S}}{\partial y}\mathbf{j} + \frac{\partial \mathrm{S}}{\partial z}\mathbf{k} \tag{A.6}$$

The magnitude of this vector represents how quickly the field is changing over the space while the direction in which the field is changing, is given just by the direction of the vector.

The divergence of a vector **a** is a scalar quantity defined as:

$$\nabla \cdot \mathbf{a} = \operatorname{div} \mathbf{a} = \frac{\partial \mathrm{a_x}}{\partial x} + \frac{\partial \mathrm{a_y}}{\partial y} + \frac{\partial \mathrm{a_z}}{\partial z} \tag{A.7}$$

Then the operator curl of vector **a** is a vector defined by:

$$\operatorname{curl} \mathbf{a} = \nabla \times \mathbf{a} = \left(\frac{\partial \mathrm{a_z}}{\partial y} - \frac{\partial \mathrm{a_y}}{\partial z}\right)\mathbf{i} + \left(\frac{\partial \mathrm{a_x}}{\partial z} - \frac{\partial \mathrm{a_z}}{\partial x}\right)\mathbf{j} + \left(\frac{\partial \mathrm{a_y}}{\partial x} - \frac{\partial \mathrm{a_x}}{\partial y}\right)\mathbf{k} \tag{A.8}$$

Finally it is possible to demonstrate that the operator *Laplacian* $\nabla^2$:

$$\nabla^2 = \frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y} + \frac{\partial^2}{\partial^2 z} \tag{A.9}$$

applied to a vector **a** is the vector given by:

$$\nabla^2 \mathbf{a} = \operatorname{grad} \operatorname{div} \mathbf{a} - \operatorname{curl} \operatorname{curl} \mathbf{a} \tag{A.10}$$

For a complete didactic explanation of Maxwell's equations and the vector operators see "A Student's Guide to Maxwell's Equations" by D. Fleisch.

2. The logarithmic units bel [bel] and decibel [dB]

Bel and decibel are logarithmic units generally used in electronics to indicate the ratio of two levels of a physical quantity like power, voltage, or current intensity.

The ratio R between two levels of electric power $W_1$ and $W_2$ is defined in bel [bel] units by:

$$\mathrm{R} = \log(\mathrm{W_1}/\mathrm{W_2}) \ \text{[bel]} \tag{A.11}$$

Much more frequently used is the unit decibel [dB] following the relation:

$$\mathrm{R} = 10 \log(\mathrm{W_1}/\mathrm{W_2}) \ \text{[dB]} \tag{A.12}$$

This ratio has the practical meaning of amplification if $W_1 > W_2$ or attenuation, with negative values, if $W_1 < W_2$. For example if $W_2$ has a reference level of 1 kW and $W_1 = 100$ kW the amplification corresponds to 20 dB. In the case of the intensity of the electric field E, the ratio is:

$$R = 10 \log (W_1/W_2) \ [\mathrm{dB}] = 20 \log (E_1/E_2) \ [\mathrm{dB}] \tag{A.13}$$

### 3. Euler's formula

In order to simplify the mathematical derivate operations, the equation in Chap. 3 Eq. 3.2.15 in the field of real numbers, $E_y = A\sin(\omega t - kx)$, can be upgraded as the Eq. 3.2.21 in the field of complex numbers, $E_y = A\exp i(\omega t - kx)$.

Note that Eqs. 3.2.15 and 3.2.21 are not equal, maintaining the link:

$$\exp i\,(\omega t - kx) = \cos\,(\omega t - kx) + i \sin\,(\omega t - kx)\,, \tag{A.14}$$

according to Euler's formula:

$$e^{is} = \cos s + i \sin s \tag{A.15}$$

where $i$ is the imaginary unit being $i^2 = -1$, and the argument $s$ expressed in radiants.

However, in Chap. 3 Eq. 3.2.15 in the field of real numbers can be derived by 3.2.21 in the field of complex numbers. In fact, the real part of the complex exponential function is able to describe the behaviour of a phenomenon like the electric field of an electromagnetic wave as:

$$\mathbf{Re}\left[\exp i\,(\omega t - kx + \varphi)\right] = \cos\,(\omega t - kx + \varphi) \tag{A.16}$$

Then also taking into account the initial phase $\varphi$, that wave can be equivalently represented by a sine or cosine function.

# Appendix B

(Additional mathematical steps to obtain the Appleton-Hartree equation)

As shown in Chap. 3, Eq. 3.2.33

$$-\mathbf{j} = i\varepsilon_0\omega X\mathbf{E} + i\mathbf{j} \times \mathbf{Y} - iZ\mathbf{j} \tag{B.1}$$

can be written along the 3 axes as:

$$-j_x = i\varepsilon_0\omega X\mathrm{E}_x - ij_zY_T - iZj_x \tag{B.2}$$

$$-j_y = i\varepsilon_0\omega XE_y + ij_zY_L - iZj_y \tag{B.3}$$

$$-j_z = i\varepsilon_0\omega XE_z + ij_xY_T - ij_yY_L - iZj_z \tag{B.4}$$

From the last two Maxwell Eqs.3.2.3 and 3.2.4 in the free space, the following relations are valid for ionospheric plasma:

$$\mathrm{curl}\,\mathbf{E} = -\delta\mathbf{B}/\delta t \tag{B.5}$$

$$\mathrm{curl}\,\mathbf{B} = \mu_0\mathbf{j} + \varepsilon_0\mu_0\,\delta\mathbf{E}/\delta t \tag{B.6}$$

where $\mathbf{j} = \sigma\mathbf{E}$ is density of electronic current with $\sigma$ the conductivity. By applying the *curl* operator to Eq. B.5:

$$curl\,\mathbf{E} = \left(\delta E_z/\delta y - \delta E_y/\delta z\right)\boldsymbol{x} + \left(\delta E_x/\delta z - \delta E_z/\delta x\right)\boldsymbol{y} + \left(\delta E_y/\delta x - \delta E_x/\delta y\right)\boldsymbol{z} \tag{B.7}$$

and taking into account that for a harmonic plane wave propagating along the $x$ axis:

$$\delta E_z/\delta y = \delta E_y/\delta z = \delta E_x/\delta z = \delta E_x/\delta y = 0; \tag{B.8}$$

the Eq. B.5gives:

$$0 = -\delta Bx/\delta t \tag{B.9}$$

$$-\delta E_z/\delta x = -\delta B_y/\delta t \tag{B.10}$$

$$\delta E_y/\delta x = -\delta B_z/\delta t \tag{B.11}$$

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1,

Now returning to the complex form of the Eq. 3.2.21 in which components of the **E** and **B** fields take the forms $E_y$=*A*exp*i(ωt−kx)* and $B_x$=*A*exp*i(ωt−kx),* the previous equations become:

$$0 = -i\omega B_x, \quad B_x = 0 \tag{B.12}$$

$$ikE_z = -i\omega B_y, \quad B_y = -kE_z/\omega \tag{B.13}$$

$$-ikE_y = -i\omega B_z, \quad B_z = kE_y/\omega \tag{B.14}$$

Similarly, applying the curl operator to the last Maxwell Eq. B.6:

$$0 = \mu_0 j_x + i\omega\,\varepsilon_0\mu_0 E_x \tag{B.15}$$

$$ikB_z = \mu_0 j_y + i\omega\,\varepsilon_0\mu_0 E_y \tag{B.16}$$

$$ikB_y = \mu_0 j_z + i\omega\,\varepsilon_0\mu_0 B_z \tag{B.17}$$

and remembering that $k=n\omega/c$:

$$j_x = -i\omega\,\varepsilon_0 E_x \tag{B.18}$$

$$ik^2 E_y/\omega = \mu_0 j_y + i\omega\,\varepsilon_0\mu_0 E_y \tag{B.19}$$

$$ik^2 E_z/\omega = \mu_0 j_z + i\omega\,\varepsilon_0\mu_0 E_z \tag{B.20}$$

$$j_x = -i\omega\,\varepsilon_0 E_x \tag{B.21}$$

$$j_y = i\omega\,\varepsilon_0\left(n^2 - 1\right)E_y \tag{B.22}$$

$$j_z = i\omega\,\varepsilon_0\left(n^2 - 1\right)E_z \tag{B.23}$$

To define the state of polarization of the wave the following two quantities are introduced: $R = E_z\,/\,E_y = -B_y\,/\,B_z$ and $S = E_x\,/\,E_y$. They produce from Eqs. B.22 and B.23 the relation $j_z = Rj_y$ and from Eqs. B.21 and B.22 the relation $j_x = S\,j_y/\,(1-n^2)$. Their substitution in Eqs. B.3 and B.4 results in:

$$S = Y_L/Y_T\left(1 + R^2\right)\left(1 - n^2\right) \tag{B.24}$$

and in Eq. B.2 results in:

$$S = i\,(Y_T/1 - X - iZ)\,R\left(1 - n^2\right) \tag{B.25}$$

By eliminating *S* from Eqs. B.24 and B.25, the solution for *R* is as follows:

$$R = \frac{i}{2\,Y_L}\left(\frac{Y_T^2}{(1 - X - iZ)} \mp \sqrt{\left(\frac{Y_T^4}{(1 - X - iZ)^2} + 4\,Y_L^2\right)}\right) \tag{B.26}$$

Now Eqs. B.3 and B.22 with the relations $j_z = Rj_y$, through the following mathematical steps:

$$-j_y = i\,\varepsilon_0\omega X E_y + iR\,j_y Y_L - iZ\,j_y \tag{B.27}$$

$$i\,\varepsilon_0\omega X E_y = -j_y - iR\,j_y Y_L + iZ\,j_{y;} \quad i\,\varepsilon_0\omega E_y = -j_y(1 + iR\,Y_L - iZ)/X, \tag{B.28}$$

give:

$$i\varepsilon_0\omega E_y = -\,i\omega\,\varepsilon_0\left(n^2 - 1\right)E_y\,(1 + iR\,Y_L - iZ)\,/X$$

or:

$$\left(n^2 - 1\right) = \frac{-\,X}{(1 + iR\,Y_L - iZ)} \tag{B.29}$$

Finally, substitution of R from Eq. B.26 produces the Appleton-Hartree equation in its standard form:

$$n^2 = (\mu - i\beta)^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1-X-iZ)} \pm \sqrt{\left(\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right)}} \tag{B.30}$$

# Bibliography

Akasofu S-I, Chapman S (1972) Solar-Terrestrial Physics. Oxford University Press, Oxford

Boschi E, Visconti G, Rizzi V, Zolesi B (1991) Il Buco dell' ozono: indizio del cambiamento globale. Editor: Il Cigno Galileo Galilei

Brekke A (1997) Physics of the Upper Atmosphere. Wiley, Praxis Publishing, Chichester

Budden KG (1961) Radio waves in the ionosphere. Cambridge University Press, Cambridge

Campbell WH (1997) Introduction to Geomagnetic Fields. Cambridge University Press, Cambridge

Davis JC (1986) Statistics and data analysis in geology. Wiley, New York

Davies K (1990) Ionospheric Radio. IEE electromagnetic waves series 31. Peter Peregrinus Ltd, London

Dominici P (1971) Radio Propagazione Ionospherica. Supplement to Vol. XXIV of Ann Geofis

Fleisch D (2008) A Student's Guide to Maxwell's Equations. Cambridge University Press, Cambridge

Handbook on ionospheric properties and propagation, ITU-Radiocommunication SG3, Geneva (1996)

Hargreaves JK (1979) The Upper Atmosphere and Solar-Terrestrial Relations. Van Nostrand Reinhold Co, England

Hargreaves JK (1992) The Solar-Terrestrial Environment. Cambridge Atmospheric and Space Science Series 5. Cambridge University Press, Cambridge

Haykin S (1994) Neural networks—a comprehensive foundation. Macmillan College Publishing Company, New York

Hunsucker RD (1991) Radio techniques for probing the terrestrial ionosphere. Springer, New York

IEEE Std 211-1997 (1998) IEEE Standard definition of terms for radio wave propagation. The Institute of Electrical and Electronics Engineers, New York

ITU-R (2001) Recommendations, radio wave propagation, vol 2000. P Series-Part 1, International Telecommunication Union, Geneva

Kaplan ED (1996) Understanding GPS Principles and Applications. Artech House, Boston

Korn GA, Korn TM (1968) Mathematical Handbook. McGraw-Hill Book, New York

Lanza R, Meloni A (2006) The Earth's Magnetism. Springer, Heidelberg

McNamara LF (1991) The Ionosphere: Communications, Surveillance, and Direction Finding. Krieger Publishing Company, Malabar

Mitra AP (1974) Ionospheric effects of solar flares. D. Reidel Publishing Company, Boston

Piggot WR, Rawer K (1972) U.R.S.I Handbook of Ionogram Interpretation and Reduction. World Data Center A for Solar-Terrestrial Physics- Report UAG-23, NOAA, Environmental Data Service, Asheville

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1,

Prölss GW (1995) Ionospheric F-region storms. In: Volland H (ed) Handbook of Atmospheric Electrodynamics. CRC Press, Boca Raton

Ractliffe JA (1962) The Magneto-Ionic Theory and its Applications to the Ionosphere. Cambridge University Press, Cambridge

Rawer K (1956) The Ionosphere. Frederick Ungar Publishing Co, New York

Rawer K (1993) Wave Propagation in the Ionosphere. Kluwer Academic Publishers, Dordrecht

Schunk RW (ed) (1996) Handbook of Ionospheric Models, Special Report, Boulder, USA, Solar-Terrestrial Energy Program (STEP), Scientific Committee on Solar Terrestrial Physics (SCOSTEP)

Sizun H (2005) Radio Wave Propagation for Telecommunication Applications. Springer, Heidelberg

Villante U (2001) Al di là delle nuvole, la fisica delle relazioni Sole-Terra. Bollati Boringhieri, Saggi Scienze

# About the Authors

**Bruno Zolesi** is a physicist involved with ionospheric and radio propagation prediction services and research since the end of the 1970s at the Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy, where he still works today. His main areas of interest and research involve the observation and investigation of the upper atmosphere, particularly the terrestrial ionosphere in relation to temporal and spatial variability, and long-term prediction and short-term forecasting models. He has a longstanding involvement in several national and international research projects, including the European COST 271 action entitled "Effects of the Upper Atmosphere on Terrestrial and Earth-space Communications", of which he was Chairman. He is a member of the International Reference Ionosphere and of the advisory board for the Journal of Space Weather and Space Climate, and was recently nominated Country Expert for the Inter-programme Coordination Team for the Space Weather of the World Meteorological Organization.

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1,

**Ljiljana R. Cander** was awarded a B.Sc in Electrical Engineering, an MSc in Plasma Physics, and a PhD in Ionospheric Research at Belgrade University, Yugoslavia. She started her scientific career at the Geomagnetic Institute of Grocka, in the areas of ionospheric measurements and aeronomy, finishing up at the Radio Communications Research Unit at the Rutherford Appleton Laboratory as Chairwoman of the European COST 296 action entitled "Mitigation of Ionospheric Effects on Radio Systems". Since her retirement in 2006, she has been a Visiting Scientist at the same Laboratory, which is now part of the UK Science and Technology Facilities Council. Her professional interests centre on ionospheric studies and applications, including propagation prediction and space weather. Dr.-Ing. Lj. R. Cander is currently Space Science Editor for Springer's Journal, Acta Geofisica, and a Fellow of the Electromagnetic Academy, USA.

# Index

B. Zolesi and L. R. Cander, *Ionospheric Prediction and Forecasting*,
Springer Geophysics, DOI: 10.1007/978-3-642-38430-1,