

Understanding Complex Systems

Springer :
COMPLEXITY

Henry D.I. Abarbanel

Predicting the Future

Completing Models of Observed
Complex Systems

 Springer

Understanding Complex Systems

Founding Editor

Prof. Dr. J.A. Scott Kelso
Center for Complex Systems & Brain Sciences
Florida Atlantic University
Boca Raton FL, USA
E-mail: kelso@walt.ccs.fau.edu

Editorial and Programme Advisory Board

Henry Abarbanel
Department of Physics and Marine Physical Laboratory (Scripps Institution of Oceanography),
University of California, San Diego, La Jolla, CA 92903, USA
Dan Braha
New England Complex Systems, Institute and University of Massachusetts, Dartmouth
Péter Érdi
Center for Complex Systems Studies, Kalamazoo College, USA and Hungarian Academy of
Sciences, Budapest, Hungary
Karl Friston
Institute of Cognitive Neuroscience, University College London, London, UK
Hermann Haken
Center of Synergetics, University of Stuttgart, Stuttgart, Germany
Viktor Jirsa
Centre National de la Recherche Scientifique (CNRS), Université de la Méditerranée, Marseille,
France
Janusz Kacprzyk
System Research, Polish Academy of Sciences, Warsaw, Poland
Kunihiko Kaneko
Research Center for Complex Systems Biology, The University of Tokyo, Tokyo, Japan
Scott Kelso
Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, USA
Markus Kirkilionis
Mathematics Institute and Centre for Complex Systems, University of Warwick, Coventry, UK
Jürgen Kurths
Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany
Andrzej Nowak
Department of Psychology, Warsaw University, Poland
Linda Reichl
Center for Complex Quantum Systems, University of Texas, Austin, USA
Peter Schuster
Theoretical Chemistry and Structural Biology, University of Vienna, Vienna, Austria
Frank Schweitzer
System Design, ETH Zürich, Zürich, Switzerland
Didier Sornette
Entrepreneurial Risk, ETH Zürich, Zürich, Switzerland
Stefan Thurner
Medical University of Vienna, Vienna, Austria

For further volumes:

<http://www.springer.com/series/5394>

Understanding Complex Systems

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition - typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels - and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems - cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The two major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Henry D.I. Abarbanel

Predicting the Future

Completing Models of Observed Complex
Systems

 Springer

Henry D.I. Abarbanel
Department of Physics and Marine Physical
Laboratory (Scripps Institution
of Oceanography)
University of California, San Diego
La Jolla, CA, USA

ISSN 1860-0832 ISSN 1860-0840 (electronic)
ISBN 978-1-4614-7217-9 ISBN 978-1-4614-7218-6 (eBook)
DOI 10.1007/978-1-4614-7218-6
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013937050

© Springer Science+Business Media New York 2013

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)

It is a pleasure to dedicate this book, and all the enjoyable hours taken in creating it, to my wife and daughters: Beth Leah Levine, Brett Lillian Levine Abarbanel, and Sara Alexis Levine Abarbanel. They tried, as a team, to morph the title into something competitive in the twenty-first century for the Amazon best sellers list. Probably they did not succeed.

Preface

Scope and Purpose of this Book

This monograph arose from my significant confusion about how to understand nervous systems as a physicist. Trying to provide a principled framework for addressing this question led me to a formulation of what we call data assimilation that has applications well beyond my initial inquiry. The point of view presented in this monograph is to view as a communications and dynamical systems problem the general challenge of transferring information in observed data to a physical (or biophysical or geophysical or ...) model of the system producing those data; this is **data assimilation**. The model, as the underlying processes, will be nonlinear in some important aspects. This problem is fundamental to how we meld experiments and observations to models, so it is hardly a new issue.

In a sense the problem is quite easy to state: over a period of time $[0, T]$ one makes observations of some properties of a physical system at some discrete times in that interval. Now we step back to think about this system and make a model of its dynamics based on past experience or intuition or whatever. This model can be expressed in differential equations or discrete time rules taking the state from one time to another.

This model could be based on physical principles and basic force laws. We would likely call this a “bottoms-up” approach to modeling the nonlinear dynamics of the processes it seeks to represent. Usually the quantities entering the model are transfer coefficients, viscosity, rate constants telling how one constituent interacts with another, thermal conductivities, etc., and, as such, usually have a distinct physical interpretation. The model could approach the problem from a “top-down” viewpoint where broad phenomenological interactions among state variables are represented by parameters with or without a more basic physical interpretation.

The methods we develop in this book do not help one create a model, except to provide information to the model about the estimated value of its parameters and a systematic method for completing the model through that estimation and to validate (or invalidate) the model through using it to predict further experiments and

observations. The methods here do not provide guidance or instruction on how to improve models that are shown to be invalid. That is still the art of the scientist. We find here a principled path to testing the consistency of a given model with given data, and that path is broadly general, useful in many application areas, and has the potential to be a tool of great use in the understanding of complex systems.

The model typically has a lot of state variables, not all of which we are able to observe, and it has some physical parameters we may not know. We want to transfer information from the measurements to that model to allow us to estimate the fixed parameters and to estimate the **unobserved** state variables in the observation interval $[0, T]$. If we can do a good job of this, then using the estimated parameters and using the estimated full model state (observed and unobserved state variables) at T , we can use the model to predict for $t > T$. To test this prediction, we require further data for $t > T$.

The prediction can go wrong for a variety of reasons: (1) The data is very noisy and the interference by the noise masks the behavior of the physical system one is trying to describe. (2) The model is wrong because it operates in a noisy environment and that masks the dynamics we want to uncover or it simply lacks dynamical elements in the differential equation or discrete time map. (3) The method used to extract information from the data and pass it along with the model is flawed; one has an insufficient data assimilation procedure. (4) The model is incorrect.

Once one has noisy data and model errors in the mix, the overall task becomes a statistical problem. There is a probability distribution for the state of the model, conditioned on the observations. One must start at $t = 0$ with an initial condition for the distribution of states. We may have some knowledge of this or may not. The idea of solving initial value problems for probabilistic quantities is totally ingrained in our physics education—think of the Schrödinger equation which produces complex probability amplitudes or the Fokker–Planck equation for real probability distributions—so this is not news. Then we must propagate this distribution function using the dynamical rules of our model to the time $0 \leq t_{\text{meas}-1} \leq T$ where we make our first measurement. At $t_{\text{meas}-1}$ we require a rule letting us know how information in the measurement at that time influences the state distribution function at that time. Then we need to propagate that distribution to the next measurement time $t_{\text{meas}-2} > t_{\text{meas}-1}$ using the dynamics and apply our information transfer rule and so forth until we reach the end of our observation window in time at T .

This set of tasks says “path integral” to contemporary physicists who have seen this question raised in quantum mechanical and statistical physics contexts over and over again. Indeed, precisely this question cast in quantum mechanical language has been a core topic in quantum theory since the 1980s (Caves 1986). The starting point for the classical version of the data assimilation problem is a path integral giving the integral representation of the solution to the data assimilation problem. The path integral is exact, as an exact statement of the information transfer at each measurement comes from an identity on conditional probabilities. The exact path integral is more or less useless for application to any specific question, so approximations and, always, numerical evaluations are required.

This book is about all of that with extensive examples from nonlinear circuits, fluid dynamics, toy geophysical models, and neurobiological simulations and experiments. The general principles are discussed both after some examples and before other examples. Indeed, we start out with a kind of standard least-squares “fitting” of a model to data and only later reveal it to be a saddle point (Laplace 1774; Debye 1909) approximation to the path integral. Since we have the integral representation to the full statistical data assimilation problem, we can formulate methods for evaluating the corrections to the saddle-point approximation, and we can formulate methods for just directly evaluating the integrals involved. The latter leads us to Monte Carlo methods and substantial optimism that contemporary parallel processing techniques will permit large problems to be solved.

The path integral is an integral representation of the linear partial differential equation for the conditional probability distribution. As such it gives the opportunity for a global view of the solution to the underlying stochastic physical problem and permits going beyond the local view of other data assimilation methodologies.

That might even permit one to investigate the stimulating questions about nervous systems as well as many other problems in complex systems.

The path integral also points our attention to the fundamental quantity in data assimilation, namely the paths of a stochastic system through its state space as they are influenced by observations. Formulating the questions one wants to answer as a path integral and focusing on performing the integrals that answer those questions bypass the efforts in other methods of data assimilation to estimate auxiliary quantities from which one might be able to extract the answers of interest. Other books give excellent instruction on those other methods, for example (Evensen 2009).

The examples presented here emphasize the importance of using the dynamical model in a series of what we call “twin experiments” to explore the requirements on measurements to perform the desired transfer of information. Twin experiments, a phrase borrowed from the geophysical literature, generate data with a given model, then uses that model with “unknown” parameters to test data assimilation methods. These twin experiments also provide a testing ground for the methods one selects for performing the path integrals at hand. Further, they are very useful for estimating the number of required measurements and identifying which measurements one needs to carry out a data assimilation task. In this, they are very helpful in designing experiments and observations.

The other theme in our formulation of statistical data assimilation is that of potential instability in the communications channel between the data as a transmitter and the model as a receiver. This is a feature of the nonlinearity of the models we use to formulate the dynamics of the underlying physical processes. It is not a feature of linear models, but it appears and has been recognized in excellent monographs (Evensen 2009). An emphasis on its importance connects the need to regularize the instabilities with the goal of using the model for predicting behavior of the physical systems of interest.

The emphasis is also on the use of prediction as the testing ground for the quality and consistency of the data, the model, and the assimilation methods. A good

“fit” to a data set by a plausible model of the underlying dynamics can be rather misleading. Stopping there and evaluating the outcome of the model representation of unobserved states is rather subjective and avoids the scrutiny a model must face in providing accurate predictions.

Returning to the nervous system questions that prompted the discussions given here, one can make any model one wishes to give a quantitative dynamical framework, and the methods in this book can be used to evaluate that model. The viewpoint on constructing quantitative, predictive models one finds here is “bottoms-up” (Rabinovich et al. 2006) starting with biophysical models and detailed experiments at the neuron level, then building the nervous system networks from that as a basis, along with further experiments, of course.

One may choose how to build a model, in this scientific arena or others, and that is an essential ingredient in using statistical data assimilation tools. The tools are indifferent to how the model is made or interpreted. It provides a path for testing those models, completed by the estimation of any unknown parameters within them.

Returning to the problem which stimulated this inquiry, the path that was initiated by the challenge of understanding functional nervous systems grew well beyond that. I made an informal survey of fields in which the data assimilation methods developed in this book are of importance and ended my survey at ten or so distinct areas ranging from toxicology and genetics to numerical weather prediction and predicting how coastal flows drive river-borne pollution dispersion. The material in this monograph might well have implications for a such a diverse set of applications, but we address only two. One is the motivating question of how one may understand from a biophysical viewpoint how functional nervous systems are constructed and operate. The other, touched upon in the chapter on twin experiments, encompasses meteorological models of the interacting atmosphere and ocean, though what is presented is only an initial study of the core ingredient for those, namely, shallow water flows. Our hope, of course, is that readers will find the material here a stepping off point for further numerical development of methods and interesting applications across many disciplines.

While perhaps a sidebar to the flow of this book, I note that I am not a fan of “punitive pedagogy” seen as the mode of presenting each idea once, and only once, regardless of its importance. The reader will find, therefore, repetition without regret presented in the hope that the pedagogy is more elevated in value.

Acknowledgments

The questions that led to the path integral formulation of data assimilation were discussed extensively with Dan Margoliash, Misha Rabinovich, Ulrich Parlitz, and Al Selverston in the context of understanding the biophysics of neurons. The confusion I noted earlier was not theirs, but mine. The ability to make the smallest steps toward understanding the issues was done in collaboration with my UCSD colleague Philip Gill. Kody J. H. Law at the University of Warwick was kind enough

to read the book in draft form and make numerous valuable suggestions for its improvement. To all of them, many thanks.

While puzzling about all this I had the opportunity to take a sabbatical from UC, San Diego, in 2009, and that provided environments where I could investigate issues without the usual distractions of home. In the laboratory of Richard Hahnloser at the University of Zurich/ETH I labored extensively with what I came to see as the saddle-point approximation to the overall answer, and at the Bernstein Center for Computational Neuroscience at the Ludwig Maximilians University in Munich hosted by Andreas Herz, I was able to formulate the first steps in understanding the questions as path integral based. The sabbatical concluded with a stay in the Margoliash lab at the University of Chicago where we began to formulate experiments that could be used to employ the path integral methods for investigating the biophysical properties of nerve cells. On returning to UCSD, I had the pleasure of many conversations with colleagues at Scripps Institution of Oceanography on these matters, especially with Bruce Cornuelle.

The research was supported by the Office of Naval Research, the Collaborative Research in Computational Neuroscience (CRCNS) jointly organized by the National Science Foundation and the National Institutes of Health, and finally by grants from the US Department of Energy, Office of Science, Applied Mathematics Program. The generous sabbatical program of the University of California allowed me to initiate and to complete the research leading to this manuscript.

This work is dedicated to my family who gave me the special context for intensely focusing on this research. I acknowledge them again here, and they can now find out what I think I was doing those many hours and months.

Contents

1	An Overview: The Challenge of Complex Systems	1
1.1	Introduction	1
1.1.1	Data Assimilation as a Communications Problem	3
1.1.2	Outline of this Book	5
2	Examples as a Guide to the Issues	7
2.1	The Malkus Waterwheel	8
2.1.1	A Physics Question About the Waterwheel	10
2.2	The Colpitts Oscillator	17
2.2.1	Colpitts Circuit Equations	19
2.2.2	Estimation with Chaotic Signals	20
2.2.3	Instability of the Synchronization Manifold	25
2.2.4	Regularized Cost Function	27
2.2.5	Experimental Colpitts Oscillator Redux	28
2.2.6	Numerical Optimization Methods	33
2.3	A Hodgkin–Huxley Neuron Model	36
2.3.1	Biophysics of the Hodgkin–Huxley Model	36
2.3.2	Estimating Parameters and Unobserved States of the HH Model	40
2.3.3	Predicting the Response of the HH Model	42
2.3.4	Consequences of the Wrong Model	45
2.4	Synopsis and Perspectives: “Slightly Complex” Examples	46
3	General Formulation of Statistical Data Assimilation	51
3.1	Data Assimilation Without Data	52
3.1.1	Deterministic Dynamics: Path Integral	52
3.1.2	Relation to the Quantum Mechanical Path Integral	56
3.1.3	Noisy Dynamics	57
3.2	Data Assimilation with a Little Bit of Data	58
3.2.1	Mutual Information	58
3.2.2	One Measurement	61
3.2.3	Two Measurements	64

3.3	The General Data Assimilation Problem	67
3.3.1	Differential Equations to Discrete Time Maps	67
3.3.2	Errors and Noise: Stochastic Data Assimilation	69
3.4	Approximating the Action	75
3.5	The Value of a Measurement	77
3.6	Predicting	78
3.7	The Scientific Value of the Path Integral	79
3.8	Data Assimilation Path Integrals in Continuous Time	80
3.9	Earlier Work on Path Integrals in Statistical Data Assimilation	83
3.10	Synopsis and Perspectives: Statistical Data Assimilation	84
4	Evaluating the Path Integral	85
4.1	Guide to Methods for Estimating the Path Integral	85
4.2	Stationary Path Methods	87
4.2.1	No Model Error	89
4.2.2	Model Errors	91
4.3	Beyond the Stationary Path: Loop Expansions	92
4.3.1	The Effective Action: Numerical Optimization Reappears	92
4.3.2	The Effective Action	93
4.3.3	Dyson–Schwinger Equations for Statistical Data Assimilation	94
4.3.4	The Effective Action: Loop Expansion	98
4.4	Estimating the Path Distribution $\exp[-A_0(\mathbf{X})]$	101
4.4.1	Langevin Equations: Fokker-Planck	101
4.5	Monte Carlo Methods	103
4.5.1	Metropolis–Hastings (Rosenbluth) Methods	103
4.5.2	Using GPU Parallel Processing	104
4.5.3	Example Monte Carlo Problem: NaKL Neuron Model	106
4.6	Consistency of Model Errors	112
4.6.1	An Example from the Lorenz96 Model with $D = 100$	115
4.6.2	An Example from the Lorenz96 Model with $D = 20$	118
4.6.3	Comments on Consistent Model Errors	122
4.7	Synopsis and Perspectives: Evaluating the Path Integral	123
5	Twin Experiments	125
5.1	The Roles of Twin Experiments	126
5.2	Neuron Models	126
5.2.1	NaKL Hodgkin–Huxley Model	127
5.2.2	The Importance of the Regularizing Variable $u(t)$	129
5.2.3	Frequency Content of the Stimulus $I_{\text{app}}(t)$	131
5.2.4	Additive Noise in the Observations	132
5.2.5	An Additional Current: I_h	134
5.2.6	NaKLh Twin Experiments	135

- 5.2.7 Model Testing Through Prediction 138
- 5.2.8 Robustness to Model Errors; NaKL Model ↔
NaKLh Model 139
- 5.2.9 NaKLh Data Presented to an NaKL Model 140
- 5.2.10 NaKL Data Driving an NaKLh Model;
Pruning a Big Model 143
- 5.3 The Lorenz96 Model 145
 - 5.3.1 Synchronization 146
 - 5.3.2 Lyapunov Exponents 149
 - 5.3.3 CLEs for Lorenz96 Models 154
 - 5.3.4 Lorenz96 Model: Variational Principle;
No Model Errors 157
- 5.4 Monte Carlo Estimation of the Path Integral
for the Lorenz96 Model 159
 - 5.4.1 How Many Observations Are Required? 162
 - 5.4.2 Results of Monte Carlo Estimation of the Path
Integral for Moments of \mathbf{X} and Model Parameters 166
 - 5.4.3 Prediction by Model Equations for $t > t_m$ 167
 - 5.4.4 Non-Gaussian Measurement Error 172
 - 5.4.5 How Often Should One Make Measurements? 172
- 5.5 Shallow Water Equations 174
 - 5.5.1 One-Layer Shallow Water Flow 175
 - 5.5.2 Statistical Data Analysis for the Shallow Water
Equations 178
 - 5.5.3 Generating the Data 180
 - 5.5.4 Synchronization of the Data with the Model Output 182
 - 5.5.5 Results for the Shallow Water Equations:
Synchronization Implies Predictability 187
- 5.6 Synopsis and Perspectives: Twin Experiments 192
- 6 Analysis of Experimental Data 199**
 - 6.1 The Avian Song System: Individual Neurons 200
 - 6.2 Experimental Procedures for HVC Neurons 202
 - 6.2.1 HVC Slice Preparation: Experimental Procedure 202
 - 6.3 Experimental Results and Analysis 203
 - 6.4 Model Details 203
 - 6.5 Details of the Numerical Evaluation 207
 - 6.5.1 Estimation and Prediction of HVC Neuron
Responses to Injected Current 208
 - 6.6 Results from Data Acquisition and Analysis 208
 - 6.6.1 Neuron 20110413_4_1 Epoch 22 208
 - 6.6.2 Neuron 20120517_1_1 Epochs 11 and 12 210
 - 6.6.3 Neuron 20120406_1_3 Epochs 19 and 15 214
 - 6.6.4 Estimated Currents and Channel Kinetics 216
 - 6.7 Comments on the Analysis of These Data 219
 - 6.8 Synopsis and Perspectives: Analysis of Experimental Data 219

7 Unfinished Business 221

 7.1 “More Measurements”: The Use of Time Delay Phase
 Space Reconstruction 223

Bibliography 227

Index 233

Chapter 1

An Overview: The Challenge of Complex Systems

1.1 Introduction

Complexity is a paradigmatic word for twenty first century science. It pervades contemporary scientific literature from regulating genetic networks to weather prediction and from hydrology to the dynamics of social networks. The sources of complex behavior as it manifests itself in the world we encounter may come from the instability of small systems expressed as irregular, nonperiodic, or chaotic space/time activity, and it may arise when one has a large numbers of degrees of freedom in the description of the fundamental physics of a system—or both, of course.

In addressing complicated trajectories in the state space of physical or other systems one is interested in a quantitative assessment of the state of the system—what does the ocean look like as seen through the lens of water flow rate or temperature—and of the ability to predict the future state of the system. Qualitative features of each, present state and predicted future state, while useful, remain insufficient, but necessary stepping stones on the way to quantitative statements.

This is certainly the view of a physicist who wishes to translate qualitative statements about interestingly rich behavior into quantitative, usually mathematical, formulations. These formulations in the hands of a physicist, and the author is certainly no exception, are not rigorous mathematical statements about the evolution equations formulated by one principle or another but are estimates, often numerical, of properties of these equations in noisy environments and in the situation where the model equations have errors. Further, measurements carrying information to the model are always noisy, so there is a degradation of properties of the observed quantities, and statistical methods are unavoidable.

Given these comments the prospect of making a science of understanding complexity sounds rather daunting, and it is. However, as new measurement instruments are developed and as enhanced computing power becomes inexpensively widely available, last year's impossibility usually becomes next year's routine. The spirit of this book is that having precise tools to bring to bear, although typically only

in approximations, to the analysis of complex systems is a productive approach to answering questions posed about these complicated systems.

The other point of view in this book is that experiments are essential in understanding and testing dynamical behavior of complex systems. Theoretical constructions alone are not going to be able to answer the set of questions we pose about such systems. While analytic solutions of simplified models of complex dynamics may continue to play an important illuminating role, we have no reason not to use available numerical tools to forge well beyond a nineteenth century fixation on analytic solutions using available numerical tools.

The challenges in translating this high level view of complex systems include

- (1) Acquiring enough accurate data from field or laboratory observations; “enough” data means observations are made with sufficient frequency in time, and at each observation time, a sufficient number of observations are performed
- (2) Developing models that embody the dynamical principles governing the source of the observations. Making models requires understanding of the physical or biological processes acting during the observations. This is quite specific to the scientific area under consideration, and we will make no general statements about this aspect of the challenges
- (3) Completing the model using information in the observations to estimate unknown model parameters and unobserved model state variables. This is one of the central arenas for this book
- (4) Examining the outcome of the estimation procedure to assure that the properties of the model correspond to one’s knowledge of the underlying dynamics. This connects with item (2) and is reliant on the model-making ability of the user
- (5) Validating the model through predicting the response of the observed system to new forcing and its time development from new initial conditions. Verification of a model, completed by the estimation of its unknown parameters and unobserved states at the time prediction begins, through prediction is the key metric for successful model building. In a philosophical sense, one can never validate a model as all possible forcing or initial states cannot be explored; however, if one is willing to work with scientific models which are “not invalid”, progress awaits

We will address many of these challenges in this book, with a strong concentration on the third and fifth items. In order to accomplish the last item, we must have accurate estimations of the model parameters and the full state of the model system when prediction begins. The overall task set here, from data collection to model validation, goes under the name of *data assimilation* in geophysical settings (Evensen 2009; Kalnay 2003; Lorenc and Payne 2007), and we adopt that name here. Further, as errors in observations, errors in the models, and uncertainty in the state space location all play a role, *statistical data assimilation* will best describe the work ahead.

1.1.1 Data Assimilation as a Communications Problem

The general issues explored in this book come from the broad scientific interest in making quantitative models of observed dynamical systems. The goal is to use the information in experimental and field observations to provide estimates for unknown parameters in the model as well as to provide an estimate of the complete state of the model system including all unobserved state variables. Systems of interest range over many orders of magnitude of time and length scales, and the interpretation of the dynamical variables in any particular scientific arena can differ completely from one area to another. When we have accomplished the estimation of the model parameters and the full model state at some time T , we can use the results to examine properties of the estimated state to determine how it associates with our physical picture of the processes involved, and we can use the model dynamical equations to predict future development of the model for $t > T$.

The scientific challenges of the present century lie in large part in the understanding of systems complex enough to elude analytic solution and large enough to elude measurements of more than a sparse subset of the dynamical variables. Large nervous systems and the dynamics of weather systems are just two examples where one seeks predictive, quantitative models. Testing these models through prediction and examining the consistency of such models with the available data is a goal of this book.

In a sense this effort is at the heart of contemporary scientific inquiry, and one can trace the quantitative feature of this scholarship back to Newton's choice of interplanetary forces once he had taken those forces to be equal to the rate of change of momentum. The hypothesis that those forces were proportional to the inverse square of the distance between two massive objects was his model whose (operational) verification came in its predictions. One's ability to solve for orbits in that force field in closed form remains a remarkable and unusual outcome of a model choice for nonlinear dynamical systems.

This problem is so pervasive that methods for accomplishing it have arisen in many communities under a variety of different names. Control theorists (Nijmeijer and Mareels 1997) call it observer theory, geoscientists (Evensen 2009; Kalnay 2003; Lorenc and Payne 2007) label it data assimilation, and many others identify it as state and parameter estimation. The core ideas and methods for accomplishing the stated goals when the dynamics of the model system are taken to be globally linear in model state space have been known for over 50 years.

The intent of this book is to address issues arising when the nonlinearity of models of physical and biophysical systems is at the fore from the outset. Everything considered here can be used in the case of globally linear dynamics. As instabilities are often at the core of the discussions, if encountered, they may signal that an assumption of global linearity is surely incorrect.

A broader view of the data assimilation task is to see it as a unidirectional communications problem. In the sense of analogies that illuminate the issues, it differs little from tuning your radio, television, smart phone, or other web device to

the broadcast signal from the source of your choice. Information in the broadcast signal may be of interest to you. It may be a song or a speech you wish to hear or a sports match you wish to view.

The transmitter at the location of the singer or the speech or the match collects audio and/or video information about the activity and modulates it onto a carrier signal which is then broadcast for you to receive. Your receiver is enabled so it can be “tuned” to the desired carrier signal through adjusting receiver characteristics such as capacitances or inductances, and then the same device is adjusted so that accurate demodulation of the information being carried can be understood at the receiving site. If the information is further encoded for efficiency of the use of transmission bandwidth or for protecting its utilization by unauthorized users, then additional “knobs” must be adjusted to allow the information to be available to the receiver.

The analogy, perhaps clear already, is that the collection of the data about a system of interest is to be compared to the production of the song or the in situ observation of the sports match. This information is packaged for transmission by measuring instruments which produce transmitted signals depending on the state of the song or match. The model, seen as a receiver, has “knobs” or parameters that need to be tuned so that the model response accurately represents the actions which were encoded by the measuring and transmitting instruments. Once the knobs/parameters are set correctly, so the transmitter and the receiver are synchronized, the output of the model will present the model builder with a clear image of the processes at work during the measurements. The communication is unidirectional as your response to the sports match, perhaps a sense of dismay if your team is losing or excitement if it is ahead, is not communicated back through the transmission channels to the players.

Of course, in the communications equipment we use on a daily basis, the transmitters and receivers have been purposely built to be tunable so they may easily and reliably be synchronized to optimize the desired communication. In the case where the information signal is from the operation of a physical system, the receiver—the dynamical model—must be built based on physical ideas and intuition, and its ability to read clearly the information in the transmission is the goal of building the model.

If the model is successful in reading one transmission with a particular setting of its knobs, and upon being sent another transmission from the same source it is, with the same setting of its knobs, able to accurately read the subsequent transmission, we have reason to accept the model as a “good” receiver. If this circumstance occurs, we may then examine the details of the model: the time dependence of its dynamical variables or of quantities dependent on those dynamical variables providing an interpretation of the processes acting to produce the data.

Further, in addition to surveying the state of the model as it represents the physical system at the source of the data, we can use that model to predict the future behavior of the dynamical system as we present it with new forcing or stimuli and provide the “gold standard” test of the quality of the model.

1.1.2 Outline of this Book

We go next to a set of examples drawn from fluid dynamics, nonlinear electrical circuits, and neurobiology. The first and third are numerical simulations; the second is accompanied by experimental data. These are selected to give a flavor of things to come. From this beginning we move to a general formulation of the statistical data assimilation problem where data are noisy, the model has errors, and the state of the model at the beginning of measurements is not known precisely. This leads directly to a formally exact solution to the overall problem in terms of a path integral through the observation window in time. The role of measurements appears as a “force” guiding the dynamical development of the model through state space. From there we turn to methods for evaluating the path integral, and these draw on the many decades of experience with the equivalent issue in statistical physics and field theory. Next we turn to a discussion of many “twin experiments” where data is generated by our model, and then the data assimilation framework is used to design forcing or stimuli that explore the dynamical dimensions of the model. These also test the methods for approximating the path integral. In a penultimate chapter we discuss the analysis of laboratory data in a neurobiological setting where the lessons from the earlier discussions are applied and realistic biophysical models are devised and tested. We recognize that the tests of these models, based on their ability to predict observed neuron voltage responses, leave us with models that are “not invalid” when prediction works. Finally, lest one suspect all is completed, we turn to a discussion of unfinished business, so far as we can identify tasks now.

Chapter 2

Examples as a Guide to the Issues

Prologue to the Chapter

Before we delve into the general structure of using information from measurements to complete models of those measurements, we will illustrate many of the questions involved by taking a look at some well-trodden ground. Completing a model means that we have estimated all the unknown parameters in the model, allowing us to predict the development of the model in its state space given a set of initial conditions and a statement of the forces acting to drive it.

We will review an example from simple fluid dynamics, the Malkus/Howard waterwheel, and then we discuss an example of a nonlinear electronic circuit, the Colpitts oscillator, widely used in commercial products. A third example will be a consideration of a Hodgkin–Huxley model of a neuron.

The first and third examples are done in computer simulation in what we call “twin experiments”. In these we generate “data” from a model with known fixed parameters. We show how we can estimate these parameters as well as the values of the unobserved and observed state variables of the dynamics at the end of an observation period, $t = T$. We then predict for times beyond T using the estimated parameters and full set of estimated state variables. In later chapters we return to the neurobiological example and use what we have learned to develop models for experimental observations of neurons from a functional neuronal network. In the second example, the Colpitts circuit, we have both numerical and experimental ingredients, yet retain substantial control over the proposed equations modeling the experiment.

The examples are quite revealing of the much harder problems we aim to address. Their formulation requires some undergraduate physics while their solution requires more demanding ideas. It is our hope that in the context of these quite simple settings, the demanding ideas stand out. The material also introduces one of the ways, a quite widely used method actually, for answering the questions of interest. We will see after the next two chapters that the approach is an approximation to the full answer to questions about estimating fixed parameters and state variables.

Though the measurements are recognized as noisy, this approach takes the models as having no errors. In that sense it is appropriate for the twin experiments covered here, but it does not provide error bars or any statistical context to the overall problem. Those important aspects will be available via the discussion upcoming in future chapters.

2.1 The Malkus Waterwheel

We begin our discussion of completing models of complex systems with examples that are only a “little bit complex”. The first example is from a fluid mechanical system. This is the waterwheel devised by Malkus in the 1970s (Malkus 1972). There are other detailed descriptions of this waterwheel primarily concentrating on the lovely idea that a simple mechanical apparatus can exhibit chaotic motions as described by Lorenz in 1963 (Strogatz 1994). An experimental study of the waterwheel is available from Illing et al. (2012a,b) who also proceed in the spirit discussed here. Their laboratory realization of the Malkus waterwheel is shown in Fig. 2.1.

The waterwheel comprises a series of leaky buckets arrayed around the circumference of a wheel of radius R . The wheel is tilted with respect to the vertical along which gravity acts. Looking at the wheel from above we associate a spatial location on the wheel by an angle θ ; $0 \leq \theta \leq 2\pi$. We describe the water in the buckets at various θ by a mass density $m(\theta, t)$, and we bring water into the buckets at a rate $Q(\theta)$. The forces on the water in the buckets drive the wheel through the torque associated with the weight of the water $g'R \int_0^{2\pi} d\theta m(\theta, t) \sin \theta$. $g' = g \sin(\alpha)$ is the effective gravitational acceleration. We select the inflow of the water to be symmetric around the top of the wheel at $\theta = 0$:

$$Q(\theta) = \sum_{n=0}^{\infty} q_n \cos(n\theta). \quad (2.1)$$

We also account for leakage from the buckets via a factor $-km(\theta, t)$ in the continuity equation expressing conservation of water. With the water source $Q(\theta)$ and the loss term, conservation of water is expressed as

$$\frac{\partial m(\theta, t)}{\partial t} + \omega(t) \frac{\partial m(\theta, t)}{\partial \theta} = Q(\theta) - km(\theta, t), \quad (2.2)$$

where the angular velocity of the wheel is $\omega(t)$. The second term on the left-hand side of this equation represents advective transport of water from one angle θ to another as the wheel rotates at angular velocity $\omega(t)$.

An equation for $\omega(t)$ comes from equating the rate of change of the angular momentum $I\omega(t)$, with I the moment of inertia of the wheel with water in the buckets, to the frictional losses in the pivot of the wheel $-\nu\omega(t)$ plus the torque on

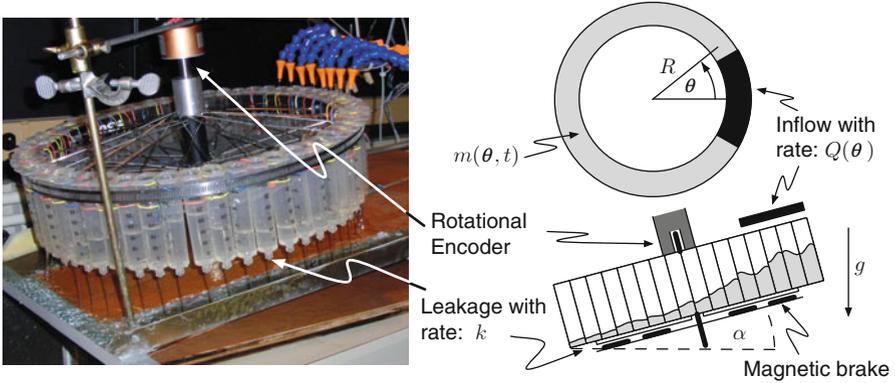


Fig. 2.1 An experimental setup of the Malkus waterwheel from Lucas Illing and his students at Reed college (Illing et al. 2012a,b). I am grateful to professor Illing for permission to use their excellent graphics. *Left* The experimental setup. *Right* Illustration of the quantities used in the Malkus dynamical equations. The **effective acceleration of gravity** in this experiment is $g \sin(\alpha)$ where α is the angle of tilt of the rotating wheel. Friction opposing the rotation is provided by a magnetic brake

the wheel arising from water in the buckets. This leads to

$$I \frac{d\omega(t)}{dt} = -v\omega(t) + g'R \int_0^{2\pi} d\theta m(\theta, t) \sin \theta. \quad (2.3)$$

Both $Q(\theta)$ and $m(\theta, t)$ are periodic in θ , so the partial differential equation for $m(\theta, t)$ may be simplified if we write

$$m(\theta, t) = \sum_{n=0}^{\infty} [a_n(t) \cos(n\theta) + b_n(t) \sin(n\theta)]. \quad (2.4)$$

Comparing coefficients of $\cos(n\theta)$ and $\sin(n\theta)$, we have

$$\begin{aligned} \frac{da_n(t)}{dt} &= -n\omega(t)b_n(t) - ka_n(t) + q_n \\ \frac{db_n(t)}{dt} &= n\omega(t)a_n(t) - kb_n(t) \\ \frac{d\omega(t)}{dt} &= -\frac{v}{I}\omega(t) + \frac{\pi g'R}{I}b_1(t). \end{aligned} \quad (2.5)$$

These equations show that knowing $\omega(t)$, $a_n(0)$, $b_n(0)$, and q_n determines $a_n(t)$ and $b_n(t)$ for all n . Also we see that the three quantities $\{a_1(t), b_1(t), \omega(t)\}$ form a separate subset of the whole infinite-dimensional space labeled by the location θ on the wheel. Solving the set of three ordinary differential equations for $n = 1$ yields $\omega(t)$ which, along with q_n , enables the determination of $m(\theta, t)$.

For $n = 0$, we have

$$\begin{aligned}\frac{da_0(t)}{dt} &= -ka_0(t) + q_0 \\ \frac{db_0(t)}{dt} &= -kb_0(t),\end{aligned}\tag{2.6}$$

indicating that $a_0(t) \rightarrow \frac{q_0}{k}$ and $b_0(t) \rightarrow 0$ exponentially rapidly in time. These then drop out of our consideration.

For $n = 1$, we have

$$\begin{aligned}\frac{da_1(t)}{dt} &= -\omega(t)b_1(t) - ka_1(t) + q_1 \\ \frac{db_1(t)}{dt} &= \omega(t)a_1(t) - kb_1(t) \\ \frac{d\omega(t)}{dt} &= -\frac{\nu}{I}\omega(t) + \frac{\pi g' R}{I}b_1(t).\end{aligned}\tag{2.7}$$

If we scale by constants $\{\alpha, \beta, \gamma, \delta\}$ and translate the dynamical variables as $a_1(t) = \alpha x(t) + \frac{q_1}{k}$, $b_1(t) = \beta y(t)$, $\omega(t) = \gamma x(t)$, and $t = \delta \tau$, we may choose the constants to arrive at

$$\begin{aligned}\frac{dx(t)}{dt} &= \sigma(y(t) - x(t)) \\ \frac{dy(t)}{dt} &= -y(t) + Ax(t) - z(t)x(t) \\ \frac{dz(t)}{dt} &= -Bz(t) + x(t)y(t) \quad B = 1.\end{aligned}\tag{2.8}$$

2.1.1 A Physics Question About the Waterwheel

Most discussions of the Malkus waterwheel (Strogatz 1994) are concerned with properties of the solutions for $a_1(t)$, $b_1(t)$, $\omega(t)$ and illustrating ideas about bifurcations and chaos in nonlinear systems. Those are important issues.

We ask here the following, somewhat different, question:

- If we observe only the rotation rate of the wheel $x_{\text{data}}(t)$ (or $\omega_{\text{data}}(t) = \frac{x_{\text{data}}(t)}{\gamma}$) over a time segment $[0, T]$, can we use the information in that time series along with a model for the waterwheel flow to determine the constants σ , A , B and the values of the unobserved state variables $y(t)$ and $z(t)$ over that period of time? Further, if we know σ , A , B and $\{x(T), y(T), z(T)\}$, can we predict the flow for $t > T$?

This is the essential kind of question we pose about complex systems where the inability to measure or observe all of the relevant variables is standard. Also present in larger and more complex problems is the chaotic motion we encounter in the first

two examples discussed here. These matters are easier to identify and diagnose in our “only slightly complex” examples, so we will attend to them for a bit. Then with the lessons in mind, we’ll turn to a general formulation of the data assimilation problem. The observation of a subset of the variables, here we observe only $\omega_{\text{data}}(t)$, is meant to represent the realistic situation we face in larger complex dynamical settings.

Returning to the simple waterwheel flow, we call estimating σ, A, B completing the model. We denote evaluating $\{x(t > T), y(t > T), z(t > T)\}$ predicting the future of our model of waterwheel flow:

- The first part completes the formulation of our model by specifying the unknown constants in the physical description of the flow.
- The second part, equally important, validates (or invalidates) our model of the flow, by utilizing our estimates of $\{x(T), y(T), z(T)\}$ as initial conditions for the model to predict the state of the flow for $t > T$.

We will demonstrate explicitly that these two tasks can be accomplished, and along the way, we will identify issues that arise even in this simple flow that make it just a “little complex”. The overall task is called *data assimilation* as information from the data is assimilated or incorporated into the model allowing prediction (Evensen 2009).

To address the question, we generated a “data set” by solving Eq. (2.8) with $\sigma_{\text{data}} = 16.0, A_{\text{data}} = 40.2, B_{\text{data}} = 1.0$ using some randomly selected initial conditions $x(0), y(0), z(0)$ within a cube in (x, y, z) space 20 “units” on a side. We integrated these equations using a standard fourth-order Runge Kutta method with a time step $\Delta t = 0.005$. The first 90,000 points of the three time series were discarded, and the remaining 2^{16} data points were recorded and stored.

We call the observed data $x_{\text{data}}(t)$, and we also know $y_{\text{data}}(t)$ and $z_{\text{data}}(t)$, we store them away for later comparison with our estimates; we treat them as unobserved variables. We inform our waterwheel flow model only of the values of $x_{\text{data}}(t)$.

This kind of exploration of data assimilation in which we generate the data from a known dynamical source and then investigate how well our methods work to produce accurate estimates of the known (unobserved) orbits and known parameters of the dynamical source is called a “twin experiment”, (Faragó et al. 2005). It is always important to exercise a twin experiment on any model of observed data.

Our method for estimating the parameters and the state variables within and at the end an data acquisition window $[0 \leq t \leq T]$ is to create a model of the data

$$\begin{aligned} \frac{dX(t)}{dt} &= \sigma(Y(t) - X(t)) \\ \frac{dY(t)}{dt} &= -Y(t) + AX(t) - Z(t)X(t) \\ \frac{dZ(t)}{dt} &= -BZ(t) + X(t)Y(t) \end{aligned} \tag{2.9}$$

and for a selection of initial conditions $X(0), Y(0), Z(0)$ and parameters σ, A, B generate a solution $X(t; X(0), Y(0), Z(0), \sigma, A, B)$ to compare to our data set $x_{\text{data}}(t)$. The comparison is made through a least-squares evaluation of the difference

$$\frac{1}{N} \sum_{n=1}^N (x_{\text{data}}(t_n) - X(t_n))^2, \quad (2.10)$$

$t_n = n\Delta t$. We then seek a minimum of this comparison metric as we vary $X(0), Y(0), Z(0)$ and σ, A, B . This choice of comparison metric assumes implicitly that the errors in the model output $X(t)$ relative to the known data $x_{\text{data}}(t)$ are Gaussian distributed.

When we try this comparison using the model we have in Eq. (2.9), this does not work well, even though we know the actual model in this twin experiment. To see the barrier to success, we fixed A, B to the values used in generating the data, fixed $X(0), Y(0), Z(0)$ to be different from that used in generating the data, and evaluated Eq. (2.10) for various values of σ . In Fig. 2.2 the green curve labeled with $K = 0$ is Eq. (2.10) with $X(t)$ directly from the model. It shows many local minima in the σ dependence of our comparison function. A search over the value of σ in producing the model output $X(t)$ is seen by eye to be very unlikely to yield the correct value ($\sigma = 16.0$) as among the many local minima; the region near $\sigma = 16.0$ would probably not be reached by a search algorithm unless we know where to start it beforehand. Further, a search algorithm is quite likely to fall into one of the numerous local minima, and usually it does.

The origin of the many local minima acting as an impediment to the estimation of σ is the incoherence between the model system output $X(t)$ and the data $x_{\text{data}}(t)$: both of them are chaotic and have different initial conditions. The instabilities in the nonlinear dynamical system that give rise to the chaos of these orbits, also lead these two orbits $X(t)$ and $x_{\text{data}}(t)$, to move differently around the same attractor of the model dynamical system with no correlation between each other as they emanate from different initial conditions, even if σ is set equal to the correct value. The two orbits $X(t)$ and $x_{\text{data}}(t)$ are not synchronized with each other.

In a manner we discuss in detail below, we can potentially tame the lack of synchronization by adding to the model equations a term that induces the model output $X(t)$ to move along with the data $x_{\text{data}}(t)$. We achieve this by changing the model equations to

$$\begin{aligned} \frac{dX(t)}{dt} &= \sigma(Y(t) - X(t)) + K(x_{\text{data}}(t) - X(t)) \\ \frac{dY(t)}{dt} &= -Y(t) + AX(t) + Z(t)X(t) \\ \frac{dZ(t)}{dt} &= -BZ(t) + X(t)Y(t); \end{aligned} \quad (2.11)$$

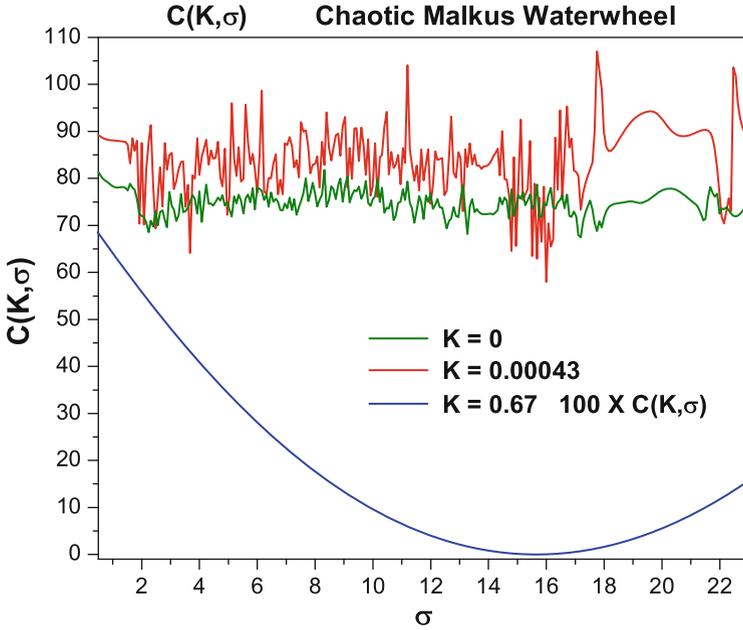


Fig. 2.2 We display the cost function $C(K, \sigma)$ Eq. (2.10) as a function of the parameter σ in the dynamical equations (2.11) for various values of K , the coupling strength of the data into the model. Other parameters in Eq. (2.11) are held fixed. The minimization of $C(K, \sigma)$ is expected to allow the estimation of σ , however, we see for small K ; $K = 0$, $K = 0.00043$ the incoherent interference of the ingredients of $C(K, \sigma)$, $x_{\text{data}}(t_n)$, and $X(t_n)$ yields many local minima with none apparently at the known value $\sigma = 16$. When the coupling of the data into the model is larger, here $K = 0.67$, the data and the model output synchronize, and the σ variation of $C(K, \sigma)$ is now smooth, allowing an accurate estimation of this parameter. The cost function $C(K, \sigma)$ decreases as $\approx K^{-2}$, so the last curve here (in blue) is multiplied by 100 so it can be displayed on an equal footing for the smaller values of K

$K \geq 0$. This connects information in $x_{\text{data}}(t)$ directly into the model equations. In the red and blue curves in Fig. 2.2 we see that when K is very small, the incoherence between $X(t)$ and $x_{\text{data}}(t)$ remains, while when we select a larger value of $K = 0.67$, the picture changes completely and the dependence of Eq. (2.10) on σ is smooth with a clear, easy to reach minimum at the correct value of σ . The key here is that the two chaotic signals $X(t)$ and $x_{\text{data}}(t)$ have synchronized for this value of K , and the previous incoherence has transmuted into coherence.

This device of adding a term to drive the two time series together is known in geophysics as “nudging” and has been used in nonlinear dynamics to analyze synchronization between chaotic systems (So et al. 1994; Parlitz 1996; Parlitz et al. 1996; Maybhate and Amritkar 1999; Sakaguchi 2002; Dochain 2003; Voss et al. 2004; Parlitz and Yu 2011; Huang 2004; Konnur 2003). It is important to note here that had we used this approach on the “ $z(t)$ ” components, it would not have worked (Pecora and Carroll 1990), a feature not recognized in the geophysics

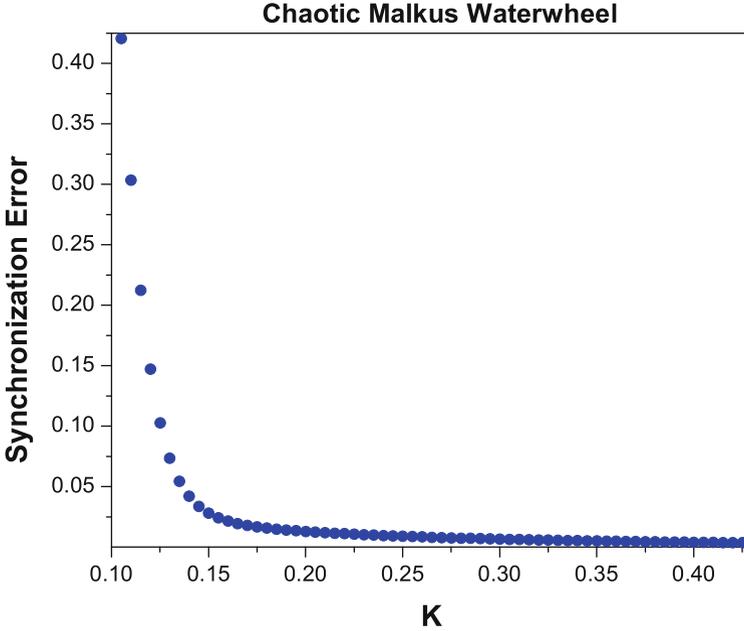


Fig. 2.3 The dependence of the synchronization error $C(K, \sigma)$ between two solutions of the dynamical equations (2.11) where all parameters are the same in the two solutions, but they differ in their initial conditions. For $K = 0$ the two chaotic solutions develop independently in time. They do not synchronize because of the instability in the “communications channel”. When the solutions are coupled as in Eq. (2.11) strongly enough to reduce the positive conditional Lyapunov exponent (CLE) at $K = 0$ below zero, the solutions synchronize, and as seen in Fig. 2.2, estimation based on minimizing the synchronization error is possible

literature, but quite relevant to our discussion as we move through this book. The new constant K we have introduced has no physical meaning, it is only a device to force the model output to the observed $x_{\text{data}}(t)$.

To orient ourselves with respect to the idea of synchronization of two chaotic time series, we plot in Fig. 2.3 a different characteristic of the function $C(K, \sigma)$:

$$C(K, \sigma) = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_{\text{data}}(t_n) - X(t_n))^2, \quad (2.12)$$

and we ask when this metric goes to zero as a function of K with σ fixed along with A and B while the initial conditions for the data and the model remain different. So, we solve the Eq. (2.11) with $\sigma = 16$ for various K and plot this least-squares metric of similarity of the model $X(t)$ time series and the data $x_{\text{data}}(t)$ times series. The outcome of this is in Fig. 2.3 where $C(K, \sigma)$ is displayed versus K ; we see a region without synchronization, $C(K, \sigma) \neq 0$, followed by persistent synchronization for larger K , $C(K, \sigma) \approx 0$.

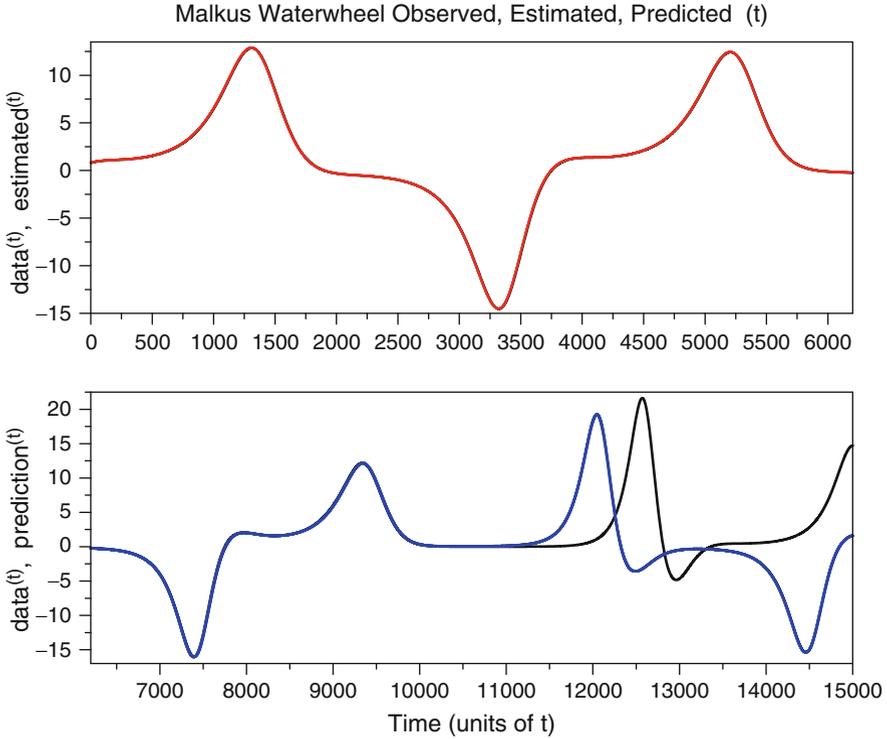


Fig. 2.4 *Top Panel* Estimation of the parameters and states for the chaotic Malkus waterwheel. Using 6,000 data points separated by $\Delta t = 0.01$ we minimized Eq. (2.13) subject to the dynamical equations (2.14). The parameters were estimated very accurately. We display the observed angular velocity comprising the data in *black* and the estimated angular velocity in *red*. The other two, unobserved, dynamical variables were also estimated throughout the observation window, but they are not displayed. *Bottom Panel* Using the estimates of the parameters and the observed and unobserved state variables at the end of the observation window, we integrated the dynamical waterwheel equations forward in time to predict the behavior of the waterwheel. The known (*black*) and predicted solutions (*blue*) agree well until about time 11,500. The estimated states and the known data at the beginning of prediction differ by small amounts, and this difference grows because of the positive Lyapunov exponent in this system. This leads to an exponential separation of the prediction and the data consistent with the value of the largest positive Lyapunov exponent

When we have synchronized the data with the appropriate function of the model output, here simply $x_{\text{data}}(t) \approx X(t)$, we expect to be more successful in our estimation of the parameters σ, A, B as well as the unobserved states $y(t), z(t)$. There is one more idea that is helpful to introduce into the discussion. In nonlinear dynamical systems such as our waterwheel equations, the instability associated with chaotic behavior of the orbits is inhomogeneous in the state space $\{X(t), Y(t), Z(t)\}$ location of the orbit (Abarbanel 1996; Kantz and Schreiber 2004). The magnitude of the instability curing factor we call K may need to be larger or smaller in various parts of the orbit. To accomplish this in a natural manner, we replace the constant

K by a function of time $K \rightarrow u(t) \geq 0$ and add a cost for the regularization accomplished by $u(t)$ to the comparison function. $u(t)$ utilizes t as an indicator where one is located in $\{X, Y, Z\}$ space and allows the strength of the push on the model toward the data to vary locally.

We then seek to minimize

$$C = \frac{1}{N} \sum_{n=1}^N \left\{ (\mathbf{x}_{\text{data}}(t_n) - X(t_n))^2 + u(t_n)^2 \right\}, \quad (2.13)$$

subject to our model equations as a constraint:

$$\begin{aligned} \frac{dX(t)}{dt} &= \sigma(Y(t) - X(t)) + u(t)(x_{\text{data}}(t) - X(t)) \\ \frac{dY(t)}{dt} &= -Y(t) + AX(t) + Z(t)X(t) \\ \frac{dZ(t)}{dt} &= -BZ(t) + X(t)Y(t), \end{aligned} \quad (2.14)$$

We have carried out this numerical optimization task, and the resulting estimates of the fixed parameters are very accurate: $\{\sigma_{\text{est}} = 16.00002, A_{\text{est}} = 40.20018, B_{\text{est}} = 0.999993\}$. In the Top Panel of Fig. 2.4 we display the estimated $\omega_{\text{est}}(t)$ in red and the known $\omega_{\text{data}}(t)$ as a black line. In addition to the estimated parameters, we estimate the unobserved variables $Y(t), Z(t)$ over the observation window $[0, T]$. This completes the model (2.9) as we now have estimates of the fixed parameters $\{\sigma, A, B\}$.

Using the estimates of σ, A, B and $X(T), Y(T), Z(T)$ we integrated the dynamical equations (2.9) with $u(t) = 0$ forward for $t > T$. In the Bottom Panel of Fig. 2.4 the predicted $\omega_{\text{pred}}(t)$ is shown in blue along with the known data still in black.

This dynamical system is chaotic, and the estimates of $\sigma, A, B, X(T), Y(T)$, and $Z(T)$ inevitably have errors, small though they may be in this example. This leads to two solutions to the waterwheel equations with slightly different initial conditions: one is the “data” here, and the other is the predicted $\omega(t)$ using the values estimated at $t = T$. As in all chaotic systems these orbits with different initial conditions diverge due to the intrinsic instability in the dynamics. This too we see in the departure of $\omega_{\text{data}}(t)$ and $\omega_{\text{pred}}(t)$ from one another (Abarbanel 1996; Kantz and Schreiber 2004).

We do not show the fact that the estimates of $\{Y(t), Z(t)\}$ are also quite accurate when we compare them to the known values in the “data” set. We could infer this by the accuracy of the predictions. In an experimental setting, not twin experiment simulations, we would not be able to make this comparison, and we would have to rely on the quality of predictions for observed state variables as our sole metric for success in estimations.

We turn now to a discussion of another “slightly complex” example drawn from nonlinear electronic circuits. We will repeat many of the steps presented in the waterwheel example and give some elaboration on the tools we utilized until now.

We did not discuss the numerical methods used to perform the minimization of the comparison function nor will we in the circuit example. We save this for the chapters following the general formulation of the data assimilation problem. In this chapter our goal is to provide a tempting sense of the flavor of the arguments and the issues and save the full meal for later.

2.2 The Colpitts Oscillator

The mechanical example of the waterwheel is not isolated or special. The issues raised in addressing data assimilation within it appear in numerous examples of nonlinear model dynamics describing observed data. To illustrate many of the same questions, we turn to an example from an electrical circuit which has a nonlinear circuit element: a bipolar transistor. In this example, we have a quite different kind of twin experiment as we generate the data from a laboratory analog circuit we built for the purpose as well as the opportunity for a fully numerical simulation of the circuit. The issues encountered in the mechanical example, which also has been realized in experiments, are now seen to be associated with the actual response of a physical system and not just a numerical simulation as before.

Our circuit is a nonlinear oscillator invented by Colpitts in the 1920s (Colpitts 1927). It has linear RLC elements, familiar from textbooks and undergraduate laboratory sessions, as well as a bipolar transistor (not part of Colpitts' original design). There are three state variables and their associated nonlinear ordinary differential equations describe this circuit; it expresses chaotic behavior (Abarbanel 1996; Kantz and Schreiber 2004; Kennedy 1994) when the forcing of the circuit becomes strong enough.

The question we propose about the circuit is whether a measured time series of one of its three independent dynamical variables can be passed to a model of the circuit and used within the model to estimate unknown parameters in the circuit model along with estimates of the remaining two unobserved model state variables. In the case of the bipolar junction transistor, we have (at least) two competing models of the nonlinear element in the circuit, and we ask if we can differentiate between them in the same setting: one measured dynamical variable passed to the model.

In practice, we are able to measure everything about the Colpitts circuit, so it provides a welcome testbed where we can examine the efficacy of any data assimilation method, and the key question is only to distinguish between two relatively well-tested models of the nonlinear bipolar transistor (Ebers and Moll 1954; Gummel and Poon 1970).

The circuit is easy enough to build (Quinn et al. 2009; De Feo and Maggio 2003), so we have the opportunity to examine it both experimentally and theoretically, and find that standard methods of data assimilation in the geophysical literature (Evensen 2009) cannot succeed when the oscillations of the Colpitts

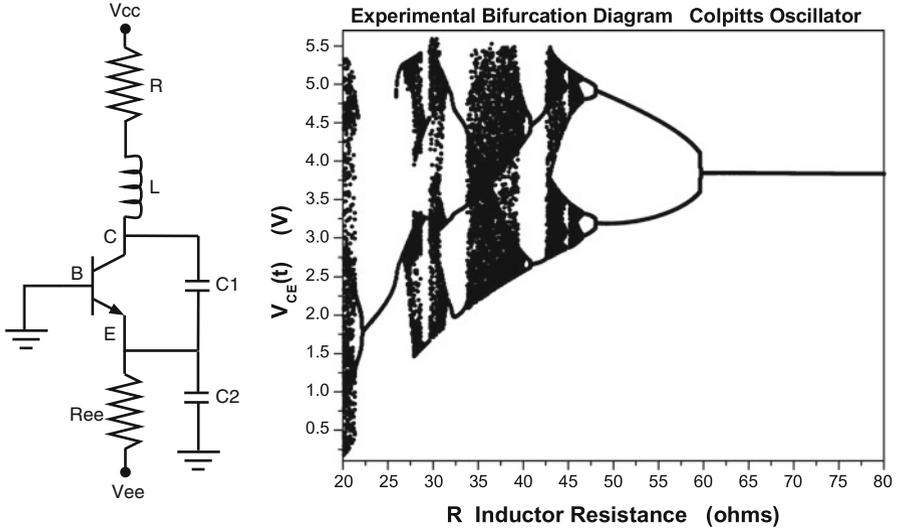


Fig. 2.5 *Left* The circuit diagram for the Colpitts circuit. The RLC circuit elements are well known from linear circuit theory (Purcell 1965), so the model of the circuit is essentially a choice of model for the bipolar transistor. *Right* The bifurcation diagram for the Colpitts circuit. As the resistance R is decreased the circuit is driven more strongly and its behavior goes from a fixed point for $R \geq 60$ ohms to periodic behavior and then chaotic oscillations for smaller values of R (Kennedy 1994). We operated this circuit with $R \approx 35$ ohms

circuit are chaotic. The techniques used to resolve this difficulty illuminate issues encountered in more complicated settings of substantial interest.

The Colpitts oscillator circuit diagram is shown in Fig. 2.5, (Left). Using the standard rules for analyzing electrical circuits (Purcell 1965), we determine there are three independent dynamical variables for the Colpitts circuit, and we choose them to be $V_{CE}(t)$, the voltage at the collector relative to the emitter; $V_E(t)$, the voltage at the emitter relative to ground; and $I_L(t)$, the current through the inductor.

Our goal in this data assimilation example is to use measurements of $V_E(t_n)$ taken at the times $t_n = n\Delta t$; $\Delta t = 10 \mu\text{s}$ for $n = 0, 1, \dots, N = 1,000$ to estimate the fixed parameters in the model of the circuit and to estimate the unobserved $V_{CE}(t_n)$ and the unobserved $I_L(t_n)$ over the same observation period. We then wish to employ the estimated fixed parameters and the estimated values of the state variables $\{V_E(T), V_{CE}(T), I_L(T)\}$ at $T = t_N = N\Delta t$ to predict the state of the system for $t > T$ using the model dynamics.

We know our estimates of the parameters and the state variables will not be perfectly accurate; therefore, if there is chaos in the system, it will exhibit itself both as inaccuracies in our prediction beyond T and as an impediment to our estimating the parameters and states of the circuit. In this section we will examine the circuit in an operating region where the driving forces are large enough to produce chaotic oscillations.

In thinking of the use of information in our data as a communications problem we would naturally seek to **synchronize** the observations with $V_E(t)$ from our model of the circuit. As the only nonlinear element of the circuit is the bipolar transistor, we can say our effort to estimate parameters and states, and predict with that information, is in essence testing our choice of a model for the physics of the transistor. In this regard, we will find that one standard model for bipolar transistors does very well in allowing estimates for states and parameters using data from this circuit, and another model does even better.

2.2.1 Colpitts Circuit Equations

The driving force in the circuit operation is the bias voltage across the circuit (Fig. 2.5 (Left)). When that driving force is small, the circuit dynamics yields a fixed point in state space. As that driving voltage is increased we encounter bifurcations (Fig. 2.5 (Right)) to limit cycle periodic oscillations of the circuit. When that voltage increases further, the circuit expresses chaotic oscillations.

The dynamics is described by three coupled first-order differential equations, which can be obtained directly using Kirchoff's laws. These are

$$\begin{aligned} C_1 \frac{dV_{CE}(t)}{dt} &= I_L(t) - I_C(V_E(t)), \\ C_2 \frac{dV_E(t)}{dt} &= I_L(t) - \frac{V_E(t) - V_{ee}}{R_{ee}} + I_B(V_E(t)), \\ L \frac{dI_L(t)}{dt} &= V_{CC} - V_E(t) - V_{CE}(t) - RI_L(t), \end{aligned} \quad (2.15)$$

where $I_C(V_E(t))$ and $I_B(V_E(t))$ are the currents going into the collector and base of the transistor.

A key ingredient is the model specifying the transistor currents. We begin with a simplified version of the Ebers–Moll equations (Ebers and Moll 1954):

$$I_C(V_E) = I_s \exp\left(\frac{-V_E}{V_T}\right), \quad (2.16)$$

$$I_B(V_E) = \frac{I_C(V_E)}{\beta_F}, \quad (2.17)$$

where β_F is the forward current gain and I_s is the reverse saturation current; both are properties of the particular transistor. $V_T = kT/e$ is the thermal voltage. These equations are nonlinear, and this is what leads to interesting behavior of the circuit.

The RLC circuit, without the transistor, has a natural period of $T_0 = 2\pi \sqrt{\frac{LC_1C_2}{C_1+C_2}}$, and has a damping time of $\tau = 2L/R$. Changing the quality factor, $Q = \pi\tau/T_0 = 2\pi L/RT_0$, can change the circuit behavior. The other parameters also matter, but

here we focus on only varying Q by changing the value of R and keeping all the other parameters fixed. As R decreases, Q increases as does the current flowing into the circuit from the terminal at V_{ee} .

The circuit was built out of common components. We used parts with the following values: $C_1 = 7.44 \mu\text{F}$, $C_2 = 7.23 \mu\text{F}$, $L = 11.74 \text{ mH}$, and $R_{ee} = 0.392 \text{ K ohms}$, and in the power supply, we took $V_{CC} = 5.03 \text{ V}$, $V_{ee} = -5.10 \text{ V}$. We also chose a 100 ohm potentiometer in the circuit to adjust the parameter R . R includes the resistance of the potentiometer plus the resistance of the inductor, which was not negligible. For the bipolar transistor we selected a 2N2222 BJT small-signal transistor. The fundamental frequency of this oscillator is around 700 Hz.

In Fig. 2.5 (**Right**) we display the bifurcation diagram for the Colpitts circuit. We show the asymptotic (in time) values taken by the voltage $V_{CE}(t)$ as a function of the resistance across the inductor. From a fixed point near $V_{CE} \approx 4 \text{ V}$ above $R \approx 60 \text{ ohm}$, we see a transition to a periodic orbit and then to windows of chaos. The chaotic oscillations used in the analysis were measured with R set at $R = 35 \text{ ohm}$. To construct this bifurcation diagram, we integrated the circuit equations using a range of values for R , and plotted the value of V_{CE} only at the times t where both $V_E(t) = V_{th}$ and $\frac{dV_E(t)}{dt} > 0$. This defines a Poincaré section (Abarbanel 1996; Kantz and Schreiber 2004; Strogatz 1994) of the three dimensional state space. The threshold value of $V_{th} = -0.6 \text{ V}$ was chosen because that is about where the transistor switches smoothly from “on” to “off”.

We recorded all three dynamical variables $V_C(t)$, $V_{CE}(t)$, and $I_L(t)$ from the circuit when it was operating in a chaotic regime. These state variables were measured every $\Delta t = 10 \mu\text{s}$. The orbit traces out a strange attractor in the three-dimensional state space. A two-dimensional projection of the experimental attractor is shown in Fig. 2.6 (**Top**). This is compared to the model projected phase space which is calculated by numerically integrating the circuit equations using parameters from the actual circuit (Fig. 2.6 (**Bottom**)).

2.2.2 Estimation with Chaotic Signals

We return to the experimental circuit in a moment. By rescaling the circuit variables, we can cast Eq. (2.15) into the dimensionless form

$$\begin{aligned} \frac{dy_1(t)}{dt} &= \alpha_D y_2(t) \\ \frac{dy_2(t)}{dt} &= -\gamma_D (y_1(t) + y_3(t)) - q_D y_2(t) \\ \frac{dy_3(t)}{dt} &= \eta_D (y_2(t) + 1 - \exp(-y_1(t))), \end{aligned} \quad (2.18)$$

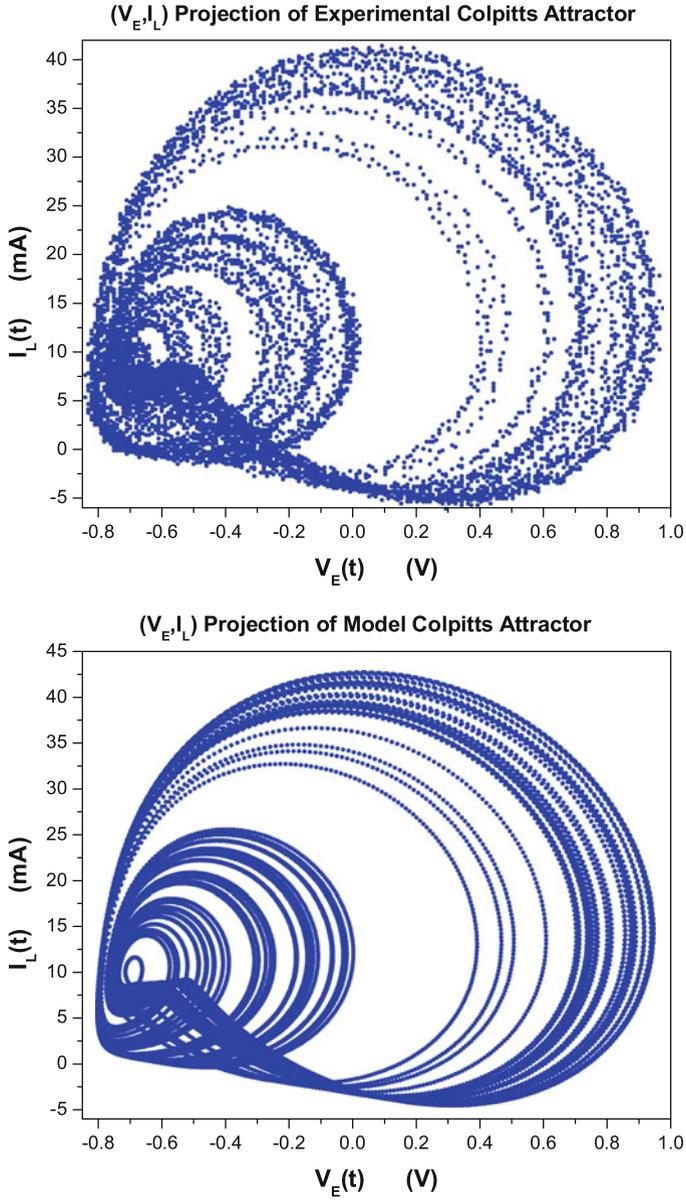


Fig. 2.6 Two-dimensional projection of the attractor: *Top Panel* Observed $V_E(t)$ versus $I_L(t)$. *Bottom Panel* $V_E(t)$ versus $I_L(t)$ from integrating the model equations. The signals were sampled every $\Delta t = 10 \mu s$

in which the control parameter α_D plays the role of the driving force. For small α_D we have a fixed point for the trajectories of the circuit. For $\alpha_D \approx 5$ we have chaotic solutions.

Now we generate data $\{y_1(t), y_2(t), y_3(t)\}$ using the parameters ($\alpha_D = 5.0$, $\gamma_D = 0.0797$, $q = 0.6898$, $\eta_D = 6.2723$) and a selected set of initial conditions $\{y_1(0), y_2(0), y_3(0)\}$. Then we ask when we can accurately estimate the parameters $\{\alpha, \gamma, q, \eta\}$ and the unobserved state variables $\{x_2(t), x_3(t)\}$ in a model of the circuit given by

$$\begin{aligned}\frac{dx_1(t)}{dt} &= \alpha x_2(t) \\ \frac{dx_2(t)}{dt} &= -\gamma(x_1(t) + x_3(t)) - qx_2(t) \\ \frac{dx_3(t)}{dt} &= \eta(x_2(t) + 1 - \exp(-x_1(t))).\end{aligned}\tag{2.19}$$

We want to perform this estimation when only the data from the state variable $y_1(t_n)$; $t_n = n\Delta t$ ($\Delta t = 10 \mu\text{s}$) is presented to the model.

If we observe $y_1(t_n) = y_1(n)$ at N points in time, we would expect that minimizing a distance between the data $y_1(n)$ and the model output $x_1(t_n) = x_1(n)$ would serve to allow estimation of the parameters $\{\alpha, \gamma, q, \eta\}$ and the initial conditions $\{x_1(0), x_2(0), x_3(0)\}$ that produce a sequence of $x_1(t_n)$; $\alpha, \gamma, q, \eta, x_1(0), x_2(0), x_3(0)$ best matched to the data $y_1(t_n)$. If the model is correct, the cost function should become quite small as the model is “tuned” to the data seen as a transmitter. In this situation, one would expect that predictions using the model equations with estimated values of the parameters and the full complement of state variables would be accurate until the chaotic nature of the orbits destabilized the predictions. If the model is incorrect, we may see a small cost function, but predictions are certain to be inaccurate. This emphasizes the role of predictions in establishing some validity for a model developed to describe experiments.

To this end, we employ a least-squares comparison distance between the data and the model output. This means we anticipate that minimizing the “cost function”

$$\frac{1}{2N} \sum_{l=0}^{N-1} (y_1(l) - x_1(l))^2,\tag{2.20}$$

subject to Eq. (2.19) should yield good estimates for parameters and state variables.

We examine a slightly different equation than Eq. (2.19) by injecting information from the data directly into the model using

$$\begin{aligned}\frac{dx_1(t)}{dt} &= \alpha x_2(t) + k(y_1(t) - x_1(t)); \quad k \geq 0 \\ \frac{dx_2(t)}{dt} &= -\gamma(x_1(t) + x_3(t)) - qx_2(t) \\ \frac{dx_3(t)}{dt} &= \eta(x_2(t) + 1 - \exp(-x_1(t))),\end{aligned}\tag{2.21}$$

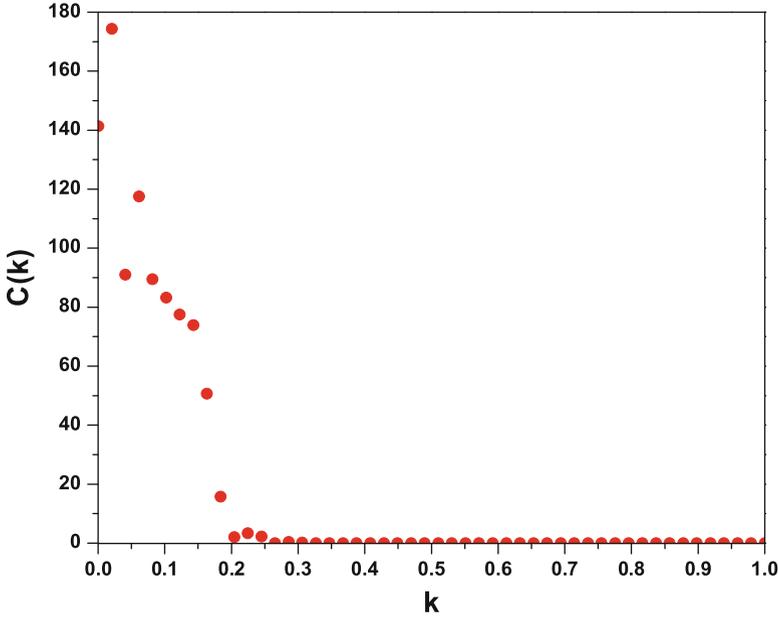


Fig. 2.7 We expect that when the data $y_1(n)$ presented to a model of the Colpitts circuit synchronizes with the corresponding output of the model $x_1(n)$, the ability to accurately estimate parameters and all states in the model will be maximized. Here we display the synchronization error $C(\eta, k) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2$ as a function of the coupling k between data and a model of the Colpitts circuit (Eq. (2.21))

which becomes Eq. (2.19) when $k = 0$. If we fix the three parameters $\{\alpha, \gamma, q\}$ and seek to estimate only η , then

$$C(\eta, k) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2 \quad (2.22)$$

should allow us to do that. The information on η and the unknown initial conditions $\{x_1(0), x_2(0), x_3(0)\}$ is contained in $x_1(t_n)$ from the solution to the model equations. The role of the coupling $k \geq 0$ is to convey information about the data $y_1(n)$ to the model equations. The term $k(y_1(t) - x_1(t))$ drives the model variable $x_1(n)$ to the data, and as $k \rightarrow \infty$, $|x_1(n) - y_1(n)| \rightarrow \frac{1}{k}$. $C(\eta, k) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2$ correspondingly decreases in magnitude as k^{-2} .

In the spirit of our overview discussion about synchronizing the data and the model, we ask for what values of k are the data $y_1(n)$ and the model output $x_1(n)$ synchronized, namely, $x_1(n) \approx y_1(n)$. For such k the cost function (2.22) should be very small. In Fig. 2.7 we show

$$C(\eta, k) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2, \quad (2.23)$$

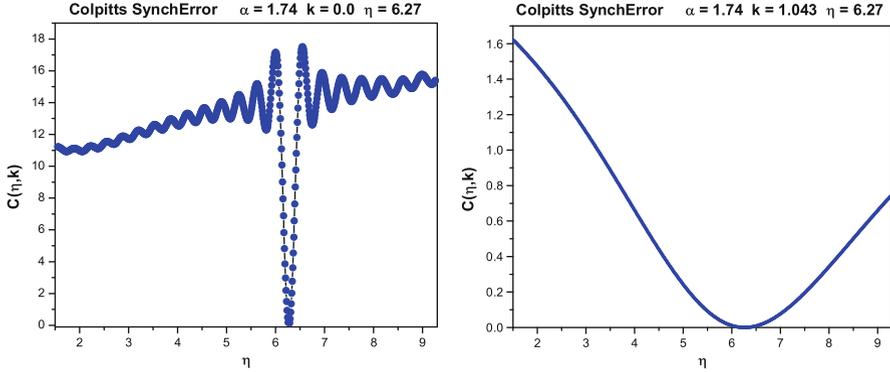


Fig. 2.8 The cost function $C(\eta, k)$ Eq. (2.23) for the Colpitts oscillator, Eq. (2.21), when the driving parameter is $\alpha = 1.74$. The circuit expresses limit cycle oscillations for this value of α . *Left* The coupling is $k = 0$. There is a global minimum at the value of the circuit parameter $\eta = 6.27$ which was used in the generation of the data. There are many local minima as well suggesting that knowledge of the correct value of η might be required to search the cost function $C(\eta, k = 0)$ for an accurate estimate. *Right* The coupling is $k = 1.043$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are synchronized. Now there is a clear minimum for the cost function $C(\eta, k = 1.043)$ allowing accurate estimation of η

as a function of k . Note that at $k = 0$ the model output and the data are not synchronized, suggesting one might not be able to estimate η or other parameters in that situation. The model and the data do synchronize for $k \geq 0.25$, and we might anticipate that when the coupling between the data stream $y_1(n)$ and the model is large enough, good estimations will occur.

To further examine this, we now look at $C(\eta, k)$ for two different values of α and two values of k : $k = 0$ and $k = 1.043$. This is one value, $k = 0$, where synchronization does not occur, and one value, $k = 1.043$, for which synchronization is present.

We first select $\alpha = 1.74$ which is in a region where a limit cycle (a periodic orbit) is present. In Fig. 2.8 (**Left**) we see that at $k = 0$ there is a distinct minimum at $\eta \approx 6.27$, though searching for it starting from an initial guess well away from the correct value might be impeded by the local minima that surrounded the correct minimum. If we increase k to $k = 1.043$, the local minima are absent (Fig. 2.8 (**Right**)), and the correct minimum where $\eta = \eta_D$ is visible and easily reached via any standard search algorithm.

When $\alpha = 5.0$ and the orbits are chaotic, we see the outcome shown in Fig. 2.9. In the **Left Panel** $k = 0$ and the two signals $y_1(n)$, the data, and $x_1(n)$ are not synchronized. This produces a ragged interference pattern where multiple local minima are present without any minimum indicating a value where $\eta = \eta_D$ can be located. However, in the right panel when $k = 1.043$, and the orbits of the data and the model output are synchronized, we see that the function $C(\eta, k)$ has become smooth, and the correct minimum is easy to locate by eye or with any sensible search algorithm.

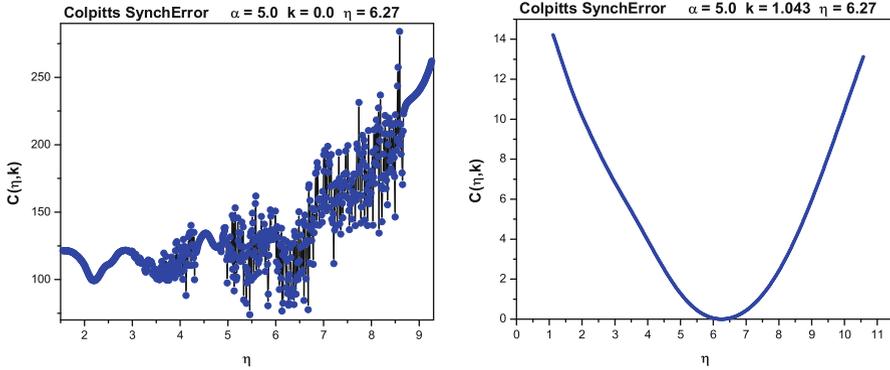


Fig. 2.9 The cost function $C(\eta, k)$, Eq. (2.23), for the Colpitts oscillator, Eq. (2.21), when the driving parameter is $\alpha = 5.0$. The circuit expresses chaotic oscillations for this value of α . *Left* The coupling is $k = 0$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are not synchronized. There are many local minima of $C(\eta, k = 0)$ none of which appear to be at the value of the circuit parameter $\eta = 6.27$ which was used in the generation of the data. *Right* The coupling is $k = 1.043$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are synchronized. Now there is a clear minimum for the cost function $C(\eta, k = 1.043)$ allowing accurate estimation of η

2.2.3 Instability of the Synchronization Manifold

Autonomous Systems: Lyapunov Exponents

How are we to understand the transition from ragged multiple local minima in $C(\eta, k)$ when $k = 0$ to the regularized, smooth $C(\eta, k)$ for larger k ? The key lies in the *conditional* Lyapunov exponents (CLEs) of the model dynamics (Pecora and Carroll 1990). These are a generalization of the usual Lyapunov exponents for a nonlinear dynamical system (Abarbanel 1996; Kantz and Schreiber 2004).

Lyapunov exponents are found by considering a discrete time $t_n = t_0 + n\Delta t$, $\mathbf{x}(t_n) = \{x_1(t_n), x_2(t_n), x_3(t_n)\}$ version of the D -dimensional dynamics, $d\mathbf{x}(t)/dt = \mathbf{F}(\mathbf{x}(t), \mathbf{p})$,

$$\mathbf{x}(t_{n+1}) = \mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(t_n), \mathbf{p}) = \mathbf{f}(\mathbf{x}(n), \mathbf{p}), \quad (2.24)$$

and asking how $\mathbf{x}(n+1)$ varies with small changes in $\mathbf{x}(0)$. The \mathbf{p} are parameters in the dynamics, and $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is the discrete time version of the model differential equation. Lyapunov exponents are usually defined for autonomous dynamics where $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is not explicitly dependent on time. In the present case this means $k(y_1(t) - x_1(t))$ is absent, so $k = 0$.

We are interested in the response of an orbit $\{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(M), \dots\}$ of the discrete time dynamics to a perturbation at t_0 . The quantity we want to analyze is the D -by- D -dimensional Jacobian matrix ($a, b = 1, 2, 3, \dots, D$)

$$\frac{\partial x_a(M+1)}{\partial x_b(0)}, \quad (2.25)$$

as this tells us quantitatively how a small change in the state of a system at t_0 leads to a change in the orbit some time later. This Jacobian matrix satisfies the recursion relation

$$\begin{aligned} \frac{\partial x_a(M+1)}{\partial x_b(0)} &= \frac{\partial f_a(\mathbf{x}(M), \mathbf{p})}{\partial x_c(M)} \frac{\partial x_c(M)}{\partial x_b(0)} \\ &= \mathbf{Df}(\mathbf{x}(M), \mathbf{p})_{ac} \frac{\partial x_c(M)}{\partial x_b(0)}. \end{aligned} \quad (2.26)$$

Repeated indices are summed over. The solution to this time-iterated map for the $D \times D$ matrix $\frac{\partial x_a(M+1)}{\partial x_b(0)}$ is

$$\begin{aligned} \frac{\partial x_a(M+1)}{\partial x_c(0)} &= \mathbf{Df}(\mathbf{x}(M), \mathbf{p})_{aa'} \mathbf{Df}(\mathbf{x}(M-1), \mathbf{p})_{a'b'} \cdots \mathbf{Df}(\mathbf{x}(1), \mathbf{p})_{b'c} \\ &= \mathbf{Df}^M(\mathbf{x}(0), \mathbf{p})_{ac}, \end{aligned} \quad (2.27)$$

noting that $\frac{\partial x_a(0)}{\partial x_b(0)} = \delta_{ab}$.

The Lyapunov exponents λ_a ; $a = 1, 2, \dots, D$ associated with the dynamics $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1)$ are given via the eigenvalues e^{λ_a} ; $\mathbf{a} = 1, 2, \dots, D$ of the Oseledec matrix (Oseledec 1968)

$$\left(\mathbf{Df}^M(\mathbf{x}(0), \mathbf{p})^T \cdot \mathbf{Df}^M(\mathbf{x}(0), \mathbf{p}) \right)^{1/2M}, \quad (2.28)$$

as $M \rightarrow \infty$. Oseledec proved that as $M \rightarrow \infty$, the λ_a

- Exist
- Are independent of $\mathbf{x}(0)$
- Are independent of the coordinate system in which the matrix is evaluated

If any of the $\lambda_a > 0$, so the eigenvalue is outside the unit circle, the system orbits are chaotic, and small perturbations to an orbit grow exponentially rapidly away from the orbit. When one or more of the $\lambda_a > 0$, orbits are unstable to small perturbations and move around the same system attractor but in a completely different order than in the unperturbed system. No orbits go off to infinity as in linear systems. Stability of the dynamical system requires $\sum_{a=1}^D \lambda_a < 0$ (Abarbanel 1996; Kantz and Schreiber 2004).

Driven Systems: Conditional Lyapunov Exponents

When $k \neq 0$, the time-iterated map is not autonomous because of the presence of $y_1(t)$ in the dynamical equations. However, if we consider the unidirectionally coupled data source and the model together as a whole system, the total dynamics

is autonomous. The Lyapunov exponents of the receiver system are called CLEs as they are conditional on the receipt of the driving signal $y_1(n)$ (Pecora and Carroll 1990). The model system is now

$$x_a(n+1) = f_a(\mathbf{x}(n), \mathbf{p}) + \delta_{a1}k(y_1(n) - x_1(n)), \quad a = 1, 2, 3, \dots, D \quad (2.29)$$

and has a Jacobian $\frac{\partial x_a(n+1)}{\partial x_b(n)} = \mathbf{Df}(\mathbf{x}(n), \mathbf{p})_{ab} - \delta_{11}k$. This Jacobian is to be used in the Oseledec formula, just as in the autonomous case, and for large k can lead to positive Lyapunov exponents becoming negative.

If any of the CLEs is positive, the synchronization manifold $\mathbf{x}(n) = \mathbf{y}(n)$ is unstable. This leads to the multiple local minima in a comparison cost function such as $C(\eta, k)$ (Eq. (2.23)). The ingredients of this cost function $x_1(n)$ and $y_1(n)$ are both chaotic but incoherent with respect to one another when unsynchronized. We see that their interference pattern, captured in the cost function, expresses multiple local minima, none of which may lie at the correct value, when a positive CLE is associated with this lack of synchronization. If one has a single unstable direction, then the role of k is to move the positive CLE to a negative value and stabilize the perturbations to the synchronization manifold introduced by the search procedure in the minimization of the cost function.

The sensitivity to small changes in initial conditions $\mathbf{x}(0)$ is also true for small changes in the parameters \mathbf{p} , and when one is searching about in $(\mathbf{p}, \mathbf{x}(0))$ space for a minimum of a cost function, these small perturbations lead, through the instability, to large excursions in the value of the cost function. This is precisely what we see in Figs. 2.8 and 2.9. As k is increased, the CLEs decrease and become negative allowing synchronization and a smooth surface in the parameter or initial condition dependence of a cost function. The fact that we can regularize the behavior of the cost function with the coupling of a single data stream $y_1(n)$ demonstrates that there is only one unstable direction on the synchronization manifold when $y_1(n)$ alone is presented to the model.

2.2.4 Regularized Cost Function

We have just seen that when the data and the model output are chaotic, synchronization does not happen when the two signals are simply compared directly. The sensitivity to initial conditions means that searching for a set of $\mathbf{x}(0)$ for the model intrinsically destabilizes the search.

The remedy presented in the previous sections adds the global coupling $k(y_1(n) - x_1(n))$ to the model dynamical equations. This fails to recognize that the instability on the synchronization manifold is not uniform over the orbits of the dynamics. Also, when the estimation procedure is completed for a fixed value of k , there is a nonzero term $k(y_1(n) - x_1(n))$ remaining in the equations. This term has no physics in it. It was introduced to rid the search procedure of sensitivity to intrinsic instabilities in the dynamics.

Both issues can be addressed by making the coupling k dependent on time, $k \rightarrow u(t)$, and by adding a cost for nonzero values of the coupling to the cost function itself. The first action recognizes that nonzero $u(t)$ maybe be needed in some parts of the system attractor, while $u(t) = 0$ may be adequate in others. The instability on the synchronization manifold is not homogeneous across the manifold. The time label in $u(t)$ serves to indicate one's location on the attractor. The second action allows us, as we achieve synchronization, to remove the time-dependent coupling as part of the variational procedure. We see this as follows:

When we take $k \rightarrow u(t)$, the equations of motion become

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t), \mathbf{p}) + \mathbf{u}(t)(\mathbf{y}(t) - \mathbf{x}(t)) ; u(t) \geq 0, \quad (2.30)$$

while we alter the cost function through

$$C(\mathbf{p}) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2 \rightarrow \frac{1}{2N} \sum_{n=0}^{N-1} [(y_1(n) - x_1(n))^2 + u(n)^2]. \quad (2.31)$$

When we minimize this cost function subject to the equations of motion, we see that as synchronization is achieved, $x_1(n) \approx y_1(n)$, the optimization procedure also sends $u(n) \rightarrow 0$.

2.2.5 Experimental Colpitts Oscillator Redux

Regularized Cost Function and Parameter Estimation

We return to the experimental setup for the Colpitts oscillator recognizing we must regulate the search procedure by using the cost function

$$C(R', u) = \frac{1}{2N} \sum_{l=1}^{N-1} [(y_1(n) - x_1(n))^2 + u(n)^2], \quad (2.32)$$

subject to the model dynamical equations

$$\begin{aligned} C_1 \frac{dV_{CE}(t)}{dt} &= I_L(t) - I_C(V_E(t)), \\ C_2 \frac{dV_E(t)}{dt} &= I_L(t) - \frac{V_E(t) - V_{ee}}{R_{ee}} + I_B(V_E(t)) + u(t)(V_{E-data}(t) - V_E(t)), \\ L \frac{dI_L(t)}{dt} &= V_{CC} - V_E(t) - V_{CE}(t) - R' I_L(t), \end{aligned} \quad (2.33)$$

and $y_1(t) = V_{E-data}(t)$, $x_1(t) = V_E(t)$.

First we examine the dependence of the cost function on the resistance R' to see if the regularization associated with stabilizing motion on the synchronization manifold is present experimentally. We coupled our experimental Colpitts oscillator to another “model” circuit, and we evaluated the cost function $C(R', u)$ as we varied the resistance R' across the inductor for four values of the coupling $u(t) = u$ in Eq. (2.33), where $u \geq 0$ is constant. We see in Fig. 2.10 the same ragged set of local minima observed in our numerical simulation of the model oscillator. As u is increased, these local minima are smoothed out, allowing for the accurate estimation of R' whose value in the data-generating circuit was 35Ω .

We estimated the unobserved state variable $V_{CE}(t)$ and $I_L(t)$ as well as the circuit parameters, when the emitter to base voltage $V_{E\text{-data}}(t) = v_1(t)$ was presented to a model of the circuit. We minimized

$$\sum_{n=1}^{1,000} [(V_{E\text{-data}}(t_n) - V_E(t_n))^2 + u(n)^2], \quad (2.34)$$

subject to the Colpitts model equations of motion, Eq. (2.33), where $t_n = n\Delta t$ and $\Delta t = 10 \mu\text{s}$ to estimate the parameters in the circuit and the values of $\{V_E(n), V_{CE}(n), I_L(n)\}$ during the observation window $0 \leq t \leq T = 10 \text{ms}$.

We also measured the parameters in the model using standard methods (Quinn et al. 2009). In Table 2.1 we show these parameters along with the estimated values when we assumed the Ebers–Moll (Ebers and Moll 1954) model for the bipolar transistor in the circuit. Except for the parameter β_F the estimates and the measured values are accurately in agreement with each other. In the estimation of the parameters shown in Table 2.2 we used the Gummel–Poon (Gummel and Poon 1970) model instead, and now we see that all the parameters are much more accurately estimated.

What should we conclude from these results of data assimilation? Well, both the Ebers–Moll and Gummel–Poon models estimate most of the parameters in the Colpitts circuit quite well. The latter repairs the one flaw in the Ebers–Moll representation of the bipolar transistor, the value of β_F , yet that model yields very good predictions. With the set of experiments we have performed, we cannot tell from the analysis of the time series alone which model to select. The Gummel–Poon model does give a slightly smaller numerical value of the cost function (Quinn et al. 2009). However, if we want to distinguish these models, we should design another set of experiments which probe the nonlinear device in other ways. One idea could be to allow the driving voltage to be time dependent in a rich manner. We will see in other examples, especially with neuron data and models, that the design of the experiment may be a critical factor in distinguishing models. Prediction will continue to be the key metric.

State Variable Estimation and Prediction

The regularized numerical optimization method yields estimates of the unobserved state variables $V_{CE}(t_n)$ and $I_L(t_n)$ in addition to estimates of the circuit parameters.

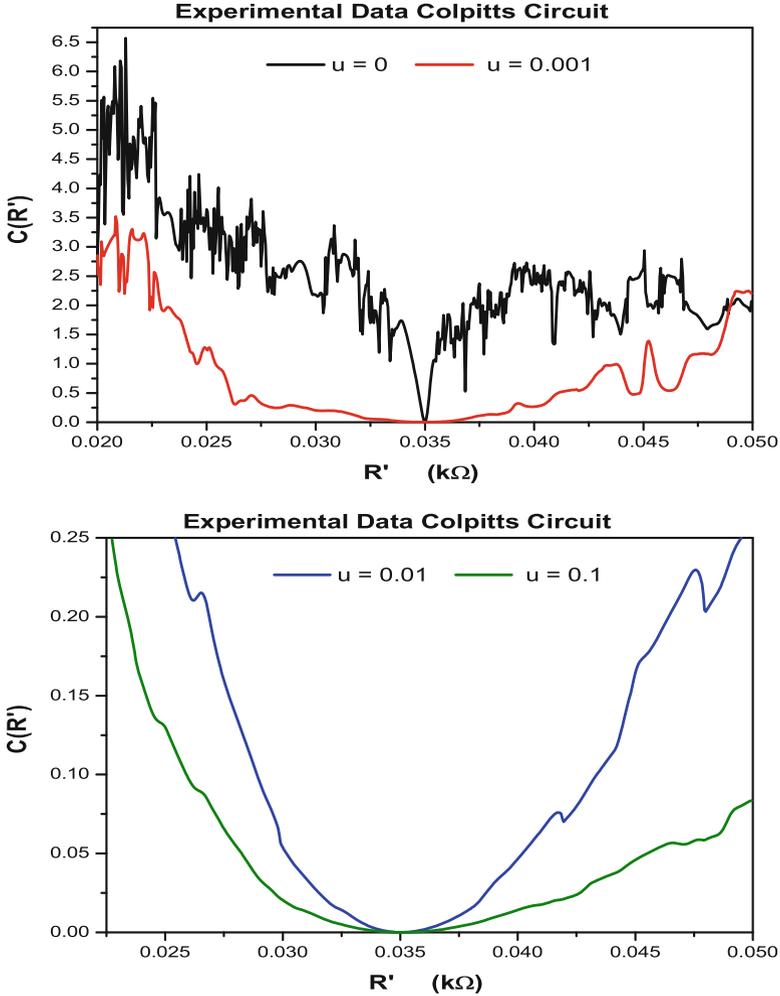


Fig. 2.10 Colpitts oscillator circuit cost function $C(R')$, Eq. (2.32), as a function of the coupling strength in the dynamical equations for the circuit Eq. (2.33). *Upper Panel* The many local minima for $u = 0$ and $u = 0.001$ arise because the voltage in the data $y_1(t_n)$ is not synchronized with the voltage $x_1(t_n)$ in the model and each is chaotic. This lack of synchronization is characterized by the CLEs: one is positive here. *Lower Panel* As we increase u to $u = 0.01$ and $u = 0.1$ the signals synchronize as the largest CLE on the synchronization manifold is moved to negative values. For the larger values of u , $C(R')$ is now smooth and the search for its minimum becomes easy. $C(R')$ decreases as $\frac{1}{u^2}$ for large u , and this effect is seen in the changed vertical scale between the upper and lower panels in the figure

In Fig. 2.11 we display the three state variables in the observation window $[0, T = 10 \text{ ms}]$. The observed values of all three state variables are displayed in red, while the values estimated using the regularized cost function are shown in black. Only $V_E(t)$ was presented to the circuit model.

Table 2.1 Estimated parameters and measured values

Name	Variational estimation	Measured	Units
C_2	7.02	7.23	μF
L	12.2	11.74	mH
R	40.0	39.3	ohms
V_0	0.661	0.63	V
V_T	25.0	27	mV
β_F	72.0	180	1

The Ebers–Moll (Ebers and Moll 1954) model of a bipolar transistor was used for the model in the minimization of the regularized cost function $C(R')$. All parameters except β_F are estimated with high accuracy

Table 2.2 Parameters estimated with the Gummel–Poon model (Gummel and Poon 1970) which includes the emitter resistance R_E along with the measured values of the parameters

Name	Variational estimation	Measured	Units
C_2	7.08	7.23	μF
L	12.00	11.74	mH
R	39.71	39.3	ohms
V_0	0.637	0.63	V
V_T	26.0	27	mV
β_F	179.0	180	1
R_E	0.23	Fixed value	ohms

Now all parameters are accurately estimated by our regularized procedure. R_E was fixed during the estimation

We may use the estimates for $\{V_E(T), V_{CE}(T), I_L(T)\}$ at the end of the observation window and the estimated circuit fixed parameters to predict the behavior of the Colpitts oscillator for $t > T$ utilizing the model differential equations, setting $u(t) = 0$. The result of this calculation is displayed in Fig. 2.12 with known values of the state variables in red and predicted values in black. Because the oscillations of the Colpitts circuit are chaotic, we expect that errors in the estimated state variables $\{V_E(T), V_{CE}(T), I_L(T)\}$ will grow in a manner associated with the positive Lyapunov exponent of the dynamics. We see that this limits the prediction horizon to about 8 ms here. This is consistent with the estimation of the positive Lyapunov exponent for the Colpitts circuit.

We set $u(t) = 0$ in making predictions with the model circuit equations as $u(t)$ is not a physical quantity but has been introduced to regularize the numerical optimization procedure. When the estimation is completed, we expect that $u(t)$ has been driven to zero by the optimization procedure and that the estimates required for prediction are independent of $u(t)$. This would tell us that the model we have used as a nonlinear filter to pass information from the observations of $V_E(t)$ to the estimates of the parameters and the unobserved state variables $\{V_{CE}(t), I_L(t)\}$ is consistent with the data. The validity of the model is examined in its ability to predict beyond the observation window.

In the estimation of states and parameters for the Colpitts model we have the values of $u(t)$ to determine whether this expectation is realized. In Fig. 2.13 we display the values of $u(t)$ in the observation window $[0, T]$, and we see the values

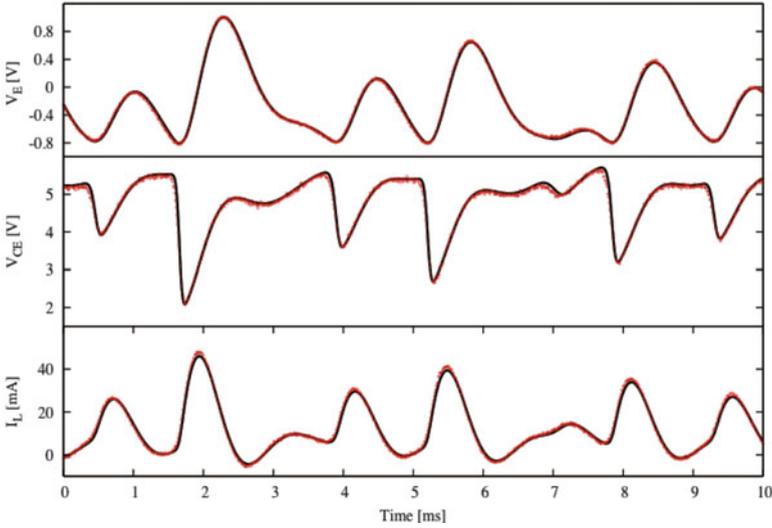


Fig. 2.11 The three Colpitts circuit dynamical variables in the observation window $0 \leq t \leq T = 10$ ms. In the *Top Panel* we see the observed $V_E(t)$ in red and the estimated values of $V_E(t)$ in black. In the *Middle Panel* we display $V_{CE}(t)$: the known value is in red, and using the regularized estimation method, we calculate the estimates of $V_{CE}(t)$ shown in black. $V_{CE}(t)$ is not presented to the model for the Colpitts oscillator. In the *Middle Panel* we display the same for $I_L(t)$: the known value is in red, and using the regularized estimation method, we calculate the estimates of $I_L(t)$ shown in black. $I_L(t)$ is not presented to the model for the Colpitts oscillator. The role of $u(t)$ is to locally regularize the surface over which we minimize the cost function (2.34). At the end of the estimation procedure, we expect $u(t) \approx 0$

appear “small”. We can also examine the smallness of the control or regulatory variable $u(t)$ in more detail by evaluating the following dimensionless ratio in

$$R(t) = \frac{F_E(t)^2}{F_E(t)^2 + [u(t)(V_{E-\text{data}}(t) - V_E(t))]^2}, \quad (2.35)$$

through the observation window. $F_E(t)$ is the portion of the vector field for the Colpitts oscillator in the equation for $\frac{dV_E(t)}{dt}$, namely,

$$F_E(t) = I_L(t) - \frac{V_E(t) - V_{ce}}{R_{ce}} + I_B(V_E(t)). \quad (2.36)$$

This tells us the relative importance of the control term $u(t)(V_{E-\text{data}}(t) - V_E(t))$ compared to the dynamics element in the equations of motion. If $R(t)$ is near unity, the role of the regularization by $u(t)$ is unimportant. We display $R(t)$ in Fig. 2.14 and note that the minimum value of $R(t)$ over the observation window is about 0.999995 in this experiment. This demonstrates that the data and the model are consistent, and the role of the control $u(t)$ is not essential. Again, the validity of the model is tested in its ability to predict.

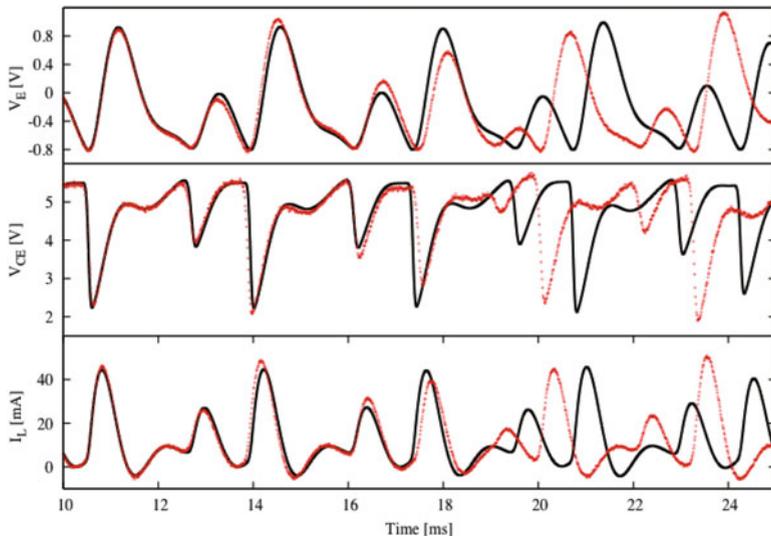


Fig. 2.12 Using the estimated values of the Colpitts state variables $\{V_E(T), V_{CE}(T), I_L(T)\}$ at $T = 10$ ms, along with the estimated parameters of the experimental circuit, we forecast the behavior of the circuit into $t > T$. The measured values of each state variable is displayed in *red*, while the predicted values are in *black*. These deviate from one another as expected because of the chaotic behavior of the circuit and the model

This ends our introduction via an example chaotic circuit realized in a laboratory setting. The estimation procedure has been to minimize a regularized cost function subject to deterministic equations of motion for the nonlinear circuit. Developing the model using Kirchoff's rules and performing the estimation of the unknown parameters and unobserved state variables are independent elements of the data assimilation task. The methods show the consistency of the model with the data and then demonstrate the ability of the completed model to predict beyond the observation window.

In an experimental setting more complex than this one we expect higher noise levels in the observations and errors in the models. Further, we anticipate that there will be uncertainty in the values of the state of the model $\mathbf{x}(t_0 = 0)$ at the start of the measurements. The use of deterministic dynamical equation constraints will not be accurate generally, and we will not turn to the development of a data assimilation framework that can account for all of these features.

2.2.6 Numerical Optimization Methods

There is no shortage whatsoever of well-developed, even very well-tested, numerical methods for solving the variational problems associated with data assimilation.

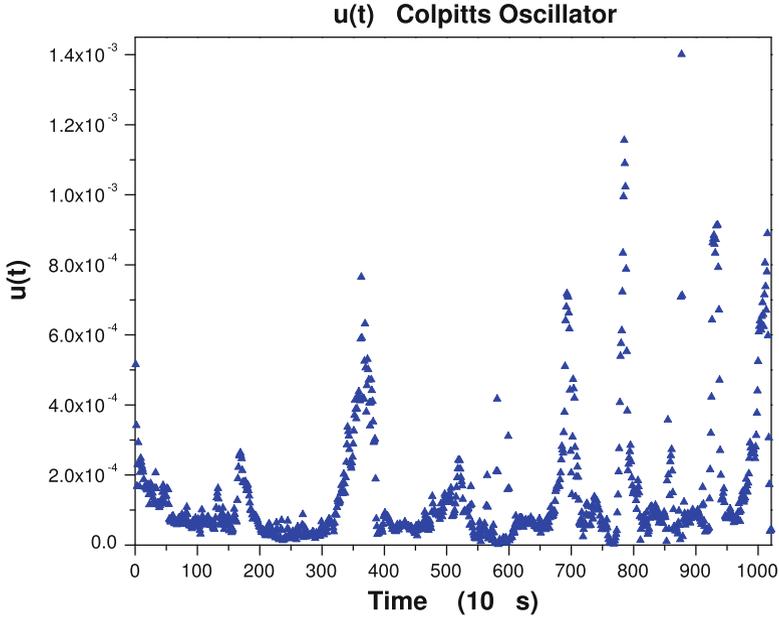


Fig. 2.13 The value of the control or regulation variable $u(t)$ throughout the observation window when the estimation procedure is completed. The dimensionless evaluation, $R(t)$, of the importance of $u(t)$ is in Fig. 2.14

Many methods use the explicit equations to find $\mathbf{x}(n > 0)$ from an initial condition $\mathbf{x}(0)$ leading to an action, often called an objective function or a cost function, which varies in a complex manner on the initial conditions $\mathbf{x}(0)$ and the parameters \mathbf{p} . In principle, one can use the same equations of motion and the estimated parameters \mathbf{p} to integrate forward from t_0 at the beginning of the observation window, where we have an estimate of $\mathbf{x}(0)$ to the end of the observation window at $t_m = T$ and beyond, $t > T$. If the model is nonlinear and has chaotic orbits in the basin of attraction of $\mathbf{x}(0)$ for the estimated parameters \mathbf{p} , the sensitivity to $\mathbf{x}(0)$ and \mathbf{p} is likely to produce quite inaccurate estimates for $\mathbf{x}(T)$ rendering prediction problematic. Indeed, estimates of $\mathbf{x}(t)$ within the observation window may suffer from this inaccuracy as well.

In the opening chapter of this book we discussed two kinds of data assimilation tasks. The first is comprised of “twin experiments” wherein one has generated data $\{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(m)\}$ from a known model and known class of measurement functions, and the purpose of exercising the data assimilation path integral, in variational approximation or otherwise, is to test one’s numerical methods. This is always important and is certainly necessary when one has new methods that estimate only $\mathbf{x}(0)$ and \mathbf{p} but might be failing. The failure is likely to come from requiring too large a time step $t_{n+1} - t_n$ or estimating states with uncertainties as to where

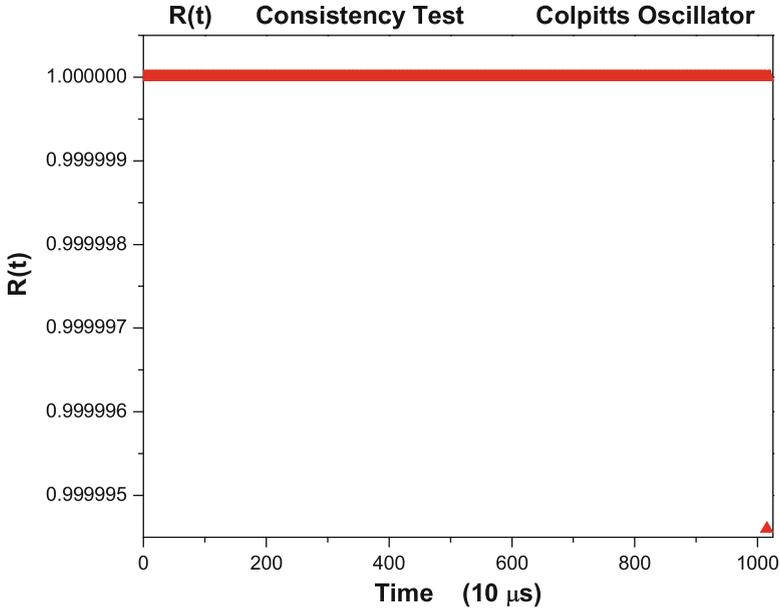


Fig. 2.14 $R(t)$, Eq. (2.35), over the observation window after the estimation procedure has completed. When $R(t)$ is near unity, the data and the model are consistent, and the role of $u(t)$ has been to regularize the search. We see here, in this twin experiment, that our model of the bipolar junction transistor is consistent with the observed experimental data

they begin (errors in $\mathbf{x}(0)$), and one can systematically correct the method for these flaws.

The other task, clearly the goal of all the methodology, is to proceed accurately when only a subset of state variables (L out of D) are known in the observation window, and estimating all of them accurately at $t_m = T$ is required. In this situation the metric of the quality of the dynamical model and its consistency with the data comes from prediction for $t > T$. If the full state $\mathbf{x}(T)$ is badly estimated, we may not know it from the data in the observation window as $D - L$ of the states are unobserved, and the L observed states may be well estimated. We will only recognize the diminished quality of the model as we predict beyond T .

This reasoning may not be important if the dynamical model has only regular behavior in state space, but nonlinear models may well have chaotic oscillations in regions of parameter and state space encountered in the search procedures. This cautions one to use the type of optimization approaches in SNOPT (Gill and Wright 1982; Gill et al. 2005) and IPOPT (Wächter and Biegler 2006).

The idea then is to use the measurements, their relation to the model state variables, and the dynamical model to establish an objective function. This is the objective function for use in the numerical packages. If the dynamics is to be used as equality constraints, namely no model errors, we seek to minimize the objective

function over the $(m + 1)D$ values of the $\mathbf{x}(n)$ plus the NP fixed parameters \mathbf{p} , subject to the mD equality constraints

$$g_a(\mathbf{x}(n + 1), \mathbf{x}(n), \mathbf{p}) = 0; \quad a = 1, 2, \dots, D; \quad n = 0, 1, \dots, m - 1. \quad (2.37)$$

This takes a problem with $D + NP$ quantities to determine and enlarges it into a problem with $(m + 1)D + NP$ quantities to vary. The space in which the numerical optimization operates can be and usually is much larger than the $D + NP$ -dimensional subspace. As explained in Gill et al. (2005) in the larger space the searches involve sparse matrices allowing for speed and accuracy. From the point of view of applying the outcome of the variational data assimilation to the understanding and prediction of the properties of the model, indeed, the testing of the model itself for consistency with the data, the apparent additional work pays off. This is called the “direct method.”

In our formulations of this strong variational problem for use with SNOPT or IPOPT, we use (Abarbanel et al. 2008) a Simpson rule integration method along with a Hermite interpolation method (Strang 1986) for implementing the equality constraints. The Simpson rule requires a midpoint evaluation, and the interpolation to the midpoint uses the Hermite method.

2.3 A Hodgkin–Huxley Neuron Model

2.3.1 Biophysics of the Hodgkin–Huxley Model

Networks of neurons exhibit rich dynamical behaviors, including rhythmic bursting and patterned sequence generation (Stein et al. 1997; Laurent et al. 2001; Johnston and Wu 1995; Koch 1999). These dynamics derive from the intrinsic properties of individual neurons and from the connections among them within the network. The membrane voltage of a neuron depends on currents from a diverse collection of ion channels, many of which have nonlinear voltage-dependent dynamics (Johnston and Wu 1995). General forms for the dynamics of many of the major families of ion channels have been characterized (ModelDB 2012; Graham 2002), but the kinetic parameters vary according to the neuron where the ion channels are located.

Neurons, from the perspective of this discussion, are nonlinear oscillators with competing feedback mechanisms associated with the flow of ions into and from a cell’s interior across channels through the cell membrane. Such a neuron model was described by Hodgkin and Huxley (Johnston and Wu 1995) some decades ago and comprises a competition between inflow of Na^+ which has a higher concentration outside than inside the cell and outflow of K^+ which has a higher concentration inside the cell. Unstimulated, a neuron maintains a resting potential relative to its exterior of about -65 mV representing a balance between diffusion of differing concentrations of ions and electrostatic forces on the charged ions. When the neuron

is forced by a positive current, its voltage rises as Na^+ ions enter the cell through a channel whose voltage-dependent permeability to Na^+ increases. As the potential passes a threshold the neuron produces a spike in voltage, called an action potential, which is tempered and turned off as the voltage-dependent permeability to K^+ ions leaving the cell rises. Action potentials travel down axons from neuron to neuron, one neuron’s output turning into another neuron’s input, providing the intercellular communication as the basis of network operation.

We look at a neuron with Na^+ and K^+ channels as well as a “leak” channel that is an adjustment factor representing leakage of charge through the cell membrane. The equations of motion comprise conservation of charge in the form

$$C \frac{dV(t)}{dt} = \text{intrinsic, voltage-dependent ion currents} \\ + \text{external stimulation currents.} \quad (2.38)$$

The intrinsic currents have the standard Hodgkin–Huxley form

$$g m(t)^{n_1} h(t)^{n_2} (E_{\text{rev}} - V(t)), \quad (2.39)$$

where n_1, n_2 are integers, E_{rev} is the Nernst potential (Johnston and Wu 1995) coming from the balance of diffusion and electrostatic forcing, $m(t)$ and $h(t)$ satisfy linear kinetic equations with voltage-dependent coefficients, and g is the maximal conductance of the ion channel. The dimensionless voltage-dependent gating variables $\{m(t), h(t)\}$ lie between 0 and 1, representing the percentage of ion channels of a particular type open or closed.

The neuron dynamics is determined with this current conservation equation along with kinetic equations for each of the $a_i(t) = \{m(t), h(t)\}$ associated with each channel:

$$\frac{da_i(t)}{dt} = \alpha_{a_i}(V(t))(1 - a_i(t)) - \beta_{a_i}(V(t))a_i(t). \quad (2.40)$$

The original HH model has four degrees of freedom or four state variables: voltage and three kinetic coefficients governing the permeability of a Na^+ channel and a K^+ channel.

Stimulating currents of different signs can reveal much of a neuron’s response properties, but due to the large variety of channels expressed in most neurons, determining the contribution of specific channel types usually requires intracellular or extracellular pharmacological manipulation to block all but the channel of direct interest. This allows the determination of the parameters in the ion current Eq. (2.39).

The efficacy and specificity of the pharmacological agents is often a concern, and data typically have to be pooled from many neurons, possibly obscuring differences among individual cells. Furthermore, such pharmacological manipulations require numerous experimental steps and lengthy recording durations, which often has

the practical implication of limiting analysis to exceptional recordings in which the neuron is fortuitously maintained in a healthy condition for an extended period of time. Further, there is no certainty that the cell membrane properties are not affected by the pharmacological agents employed. Based on these results, biophysical Hodgkin–Huxley models of the neurons can be constructed, with the parameters fixed by experimental results or chosen based on simulation and iterative search (ModelDB 2012).

Our approach to the question of estimating the parameters and states of an HH neuron works directly with the measured $V(t)$ and the selected stimulating current. Pharmacology is bypassed.

We consider the most common case, where the only measurement is the membrane voltage when a selected stimulus current is applied to the cell. Measuring a time series for $a_i(t)$ directly is not presently possible. A stimulus current $I_{\text{app}}(t)$ is applied through the same electrode recording the membrane voltage to probe the responses of voltage-gated channels. A data set usually consists of the applied current and the observed voltage over some observation window $t = [t_0, t_m = T]$ with measurements made at times $t_n = \{t_0, t_1, \dots, t_m\}$ within that window.

Working solely with observations of the membrane voltage and a known stimulus or driving function of the neuron, the goal is to estimate all of the fixed parameters of the model, including maximal conductance, intracellular ion concentrations, and channel-gating kinetics as well as the unobserved gating dynamics $a_i(t)$ throughout the observation window. Knowledge of the quantities $V(T)$ and all of the $a_i(T)$ along with estimates of all fixed parameters in the full model allow us to use the model to predict the response of the neuron for $t > T$ given $I_{\text{app}}(t > T)$. If these predictions are accurate for a broad range of biologically plausible stimuli, then the estimates of the model parameters provide a parsimonious, biophysically interpretable description of the neuron’s behavior. Furthermore, such results motivate the hypothesis that the model neurons will respond accurately to a diverse set of stimuli when used in models of interesting functional networks. This suggests that networks of such neurons could be constructed in a two-stage process: (1) using stimulus, response data to analyze the classes of individual neurons in the network using the dynamical methods outlined here, and (2) determine the connections among such neurons recognizing the connection strengths as parameters in a dynamical system using data on the network as a whole.

These requirements raise a number of practical and theoretical questions:

1. How many measurements $L \leq D$ are required? Typically for single neurons $L = 1$ (voltage), and given that most of the states of the system will remain unmeasurable, $L \ll D$. Equivalently, given that observations are generally limited to $V(t)$, is there an upper limit to the complexity of the models that can be completed from the data?
2. How often is it necessary to make measurements? Namely, given a fixed observation window $[t_0, \dots, t_m = T]$, how frequently should the measurable variables be sampled? How large can one choose the intervals between observations, $t_{n+1} - t_n$?

3. Given that measurements are noisy and models are always wrong in some respect, how robust are the estimates of parameters and unobserved state variables to these errors?
4. What kinds of stimuli lead to adequate exploration of the state space of the model so that all the parameters and unobserved state variables can be estimated with a similar degree of confidence?
5. What metrics are most appropriate for testing the validity of a neuronal model, especially in light of the intrinsic variability of real neurons?

The model we address (Abarbanel et al. 2011; Toth et al. 2011) consists of the dynamics of membrane voltage $V(t)$ driven by two voltage-gated ion channels, Na and K, and a “leak” current and an external, applied current we call $I_{\text{app}}(t)$. The dynamics of the model comprise the equations for voltage

$$\begin{aligned} \frac{dV(t)}{dt} &= \frac{1}{C} \{ g_{\text{Na}} m(t)^3 h(t) (E_{\text{Na}} - V(t)) + g_{\text{K}} n(t)^4 (E_{\text{K}} - V(t)) \\ &\quad + g_{\text{L}} (E_{\text{L}} - V(t)) + I_{\text{DC}} + I_{\text{app}}(t) \}, \\ &= F_V(V(t), m(t), h(t), n(t)) \end{aligned} \quad (2.41)$$

where the g_X 's indicate maximal conductances and the E_X 's reversal potentials, for each of the Na, K, and leak channels. I_{DC} is a DC current, and $I_{\text{app}}(t)$ is an applied time-dependent external current selected in an experiment. We refer to this as the NaKL HH model.

The gating variables $a_i(t) = \{m(t), h(t), n(t)\}$ are discussed in many textbooks and reviews (Johnston and Wu 1995; ModelDB 2012; Graham 2002), and each satisfies a first-order kinetic equation of the form

$$\frac{da_i(t)}{dt} = \frac{a_{i0}(V(t)) - a_i(t)}{\tau_i(V(t))}. \quad (2.42)$$

The kinetic terms $a_{i0}(V)$ and $\tau_i(V)$ are taken here in the form

$$\begin{aligned} a_{i0}(V) &= \frac{1}{2} \left(1 + \tanh \left[\frac{(V - v_a)}{dva} \right] \right) \\ \tau_i(V) &= t_{a0} + t_{a1} \left(1 - \tanh^2 \left[\frac{(V - vat)}{dvat} \right] \right) \text{ or} \\ \tau_i(V) &= t_{a0} + t_{a1} \tanh \left[\frac{(V - vat)}{dvat} \right]. \end{aligned} \quad (2.43)$$

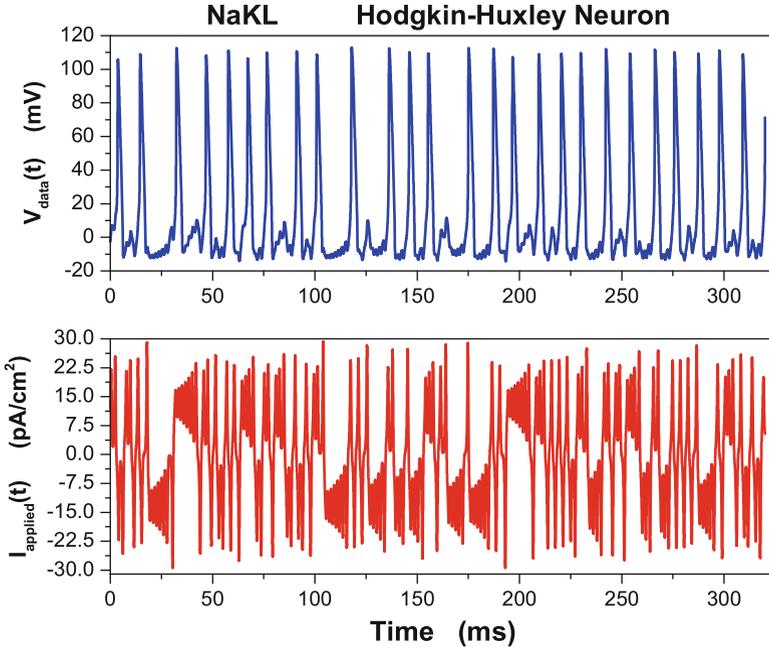


Fig. 2.15 NaKL Hodgkin–Huxley neuron model. *Top Panel* The “data” in this twin experiment. The response membrane voltage. *Bottom Panel* The stimulating current. This is an **observed** state variable

In the language used in our general formulation, the model state variables are $\mathbf{x}(t) = \{V(t), m(t), h(t), n(t)\}$, and the parameters are $\mathbf{p} = \{C, g_{\text{Na}}, E_{\text{Na}}, g_{\text{K}}, E_{\text{K}}, \dots, \text{dvat}\}$. In a twin experiment, the data $\{V_{\text{data}}(t), m_{\text{data}}(t), h_{\text{data}}(t), n_{\text{data}}(t)\}$ are generated by solving these HH equations for some initial condition $\mathbf{x}(0) = \{V(0), m(0), h(0), n(0)\}$ and some choice for the parameters, the DC current and $I_{\text{app}}(t)$, and the stimulating current. The data $y(t)$ presented to the model consist of only the voltage $V_{\text{data}}(t)$ along with additive noise: $y(t) = V_{\text{data}}(t) + \text{noise}(t)$.

2.3.2 *Estimating Parameters and Unobserved States of the HH Model*

We solved the HH equations (2.41) and (2.42) using the stimulating current shown in the Bottom Panel of Fig. 2.15. The Top Panel of this figure displays the time series of $V(t)$ resulting from the calculation. A fourth-order Runge Kutta method was employed using a time step of 0.01 ms. This current was selected to probe the dynamical range of the currents in the NaKL model including the spiking region for high voltage and the subthreshold region. The current wave form was taken from

Table 2.3 Parameters used in the NaKL HH model to generate data for a twin experiment

Name	Value	Name	Value
C	$1.0 \mu\text{F}/\text{cm}^2$	vh	-60.0 mV
g_{Na}	$120.0 \text{ mS}/\text{cm}^2$	dvh	-15.0 mV
E_{Na}	50.0 mV	t_{h0}	1.0 ms
g_{K}	$20.0 \text{ mS}/\text{cm}^2$	t_{h1}	7.0 ms
E_{K}	-77.0 mV	vht	-60.0 mV
g_{L}	$0.3 \text{ mS}/\text{cm}^2$	dvht	-15.0 mV
E_{L}	-54.4 mV	vn	-55.0 mV
vm	-40.0 mV	dvn	30.0 mV
dvm	15.0 mV	t_{n0}	1.0 ms
t_{m0}	0.1 ms	t_{n1}	5.0 ms
t_{m1}	0.4 ms	vnt	-55.0 mV
vmt	-40.0 mV	dvnt	30.0 mV
dvmt	15.0 mV	I_{DC}	$7.3 \text{ pA}/\text{cm}^2$

The model includes Na, K, and leak currents

the output of the Malkus waterwheel, but its essential ingredients are that it changes slowly compared to the response of the neuron and causes the neuron response to visit significant parts of the $\{V(t), m(t), h(t), n(t)\}$ state space. To the equations of motion for $V(t)$, we added a regulating term $u(t)(y(t) - V(t))$.

When the cost function

$$\frac{1}{N} \sum_{n=0}^{N-1} \left[(y(t_n) - V(t_n))^2 + u(t_n)^2 \right] \quad (2.44)$$

was minimized subject to Eqs.(2.41) and (2.42), we determined the 18 fixed parameters in the model along with the values of $\{V(t), m(t), h(t), n(t)\}$ at each point in the observation window $\{t_0 = 0, \dots, t_n = n\Delta t, \dots, t_m = T = 90 \text{ ms}\}$. This involves a numerical minimization in 36,018 dimensional space. This was accomplished in order of ten minutes on a commodity PC.

The parameters selected and the outcome of the numerical optimization are shown in Tables 2.3 and 2.4. In Fig.2.16, we show in the **Left Panel** the known, “data”, membrane voltage along with the estimated $V(t)$ resulting from the optimization procedure. In the **Right Panel** we display the known Na^+ activation gating variable $m(t)$ along with the estimation of this resulting from the numerical optimization. $m(t)$ is an **unobserved** state variable. Its estimation is achieved by the nonlinear filtering performed by the dynamical equations of motion used as constraints on the cost function as it is minimized. In treating experimental data, $m(t)$ and any other gating variables are not observed, and we have no way to examine the precision with which they are estimated. One of the important uses of twin experiments is that it permits us to assess how well unobserved quantities are being estimated and, in that manner, build confidence in our overall state and parameter estimation effort.

Table 2.4 Parameters in generating the “Data” and the estimated parameters in NaKL HH model; I_{DC} was fixed in all calculations

Name	Value in “data”	Estimated value
g_{Na} (mS/cm ²)	120	118.79
E_{Na} (mV)	50.0	49.99
g_K (mS/cm ²)	20	20.35
E_K (mV)	-77.0	-76.97
g_L (mS/cm ²)	0.3	0.2955
E_L (mV)	-54.4	-54.11
$v_m = v_{mt}$ (mV)	-40.0	-40.08
$d_v m = d_{vmt}$ (mV)	15.0	14.90
t_{m0} (ms)	0.1	0.1009
t_{m1} (ms)	0.4	0.3982
$v_h = v_{ht}$ (mV)	-60.0	-59.91
$d_{v h} = d_{vht}$ (mV)	-15.0	-14.88
t_{h0} (ms)	1.0	1.004
t_{h1} (ms)	7.0	7.047
$v_n = v_{nt}$ (mV)	-55.0	-54.91
$d_{v n} = d_{vnt}$ (mV)	30.0	29.28
t_{n0} (ms)	1.0	1.012
t_{n1} (ms)	5.0	4.99

As the term $u(t)(y(t) - V(t))$ introduced the nonphysical terms $u(t_n)$ into the cost function, we would like to see that at the end of the numerical procedures it is absent. To assess this we evaluated the dimensionless quantity

$$R(t) = \frac{F_V(V(t), m(t), h(t), n(t))^2}{F_V(V(t), m(t), h(t), n(t))^2 + (u(t)(y(t) - V(t)))^2} \quad (2.45)$$

throughout the observation window using the final values for all quantities at the termination of the estimation procedure. This is shown in Fig. 2.17. We see that $R(t)$ is effectively united for all times in the observation window, showing the unimportance of the unphysical control term at the final stage of the numerical process.

2.3.3 Predicting the Response of the HH Model

A numerical optimization method solving the constrained optimization problem “minimize this cost function subject to these equations as constraints” will always produce numbers. In the case of twin experiments we can test the accuracy of those numbers as we know them a priori. In the more interesting case of experiments where only the stimulating currents and the response voltages are known to us, we need to ask more of the numerical output.

We take as the primary goal of the data assimilation effort we describe through this monograph the prediction of new response voltages when the model neuron

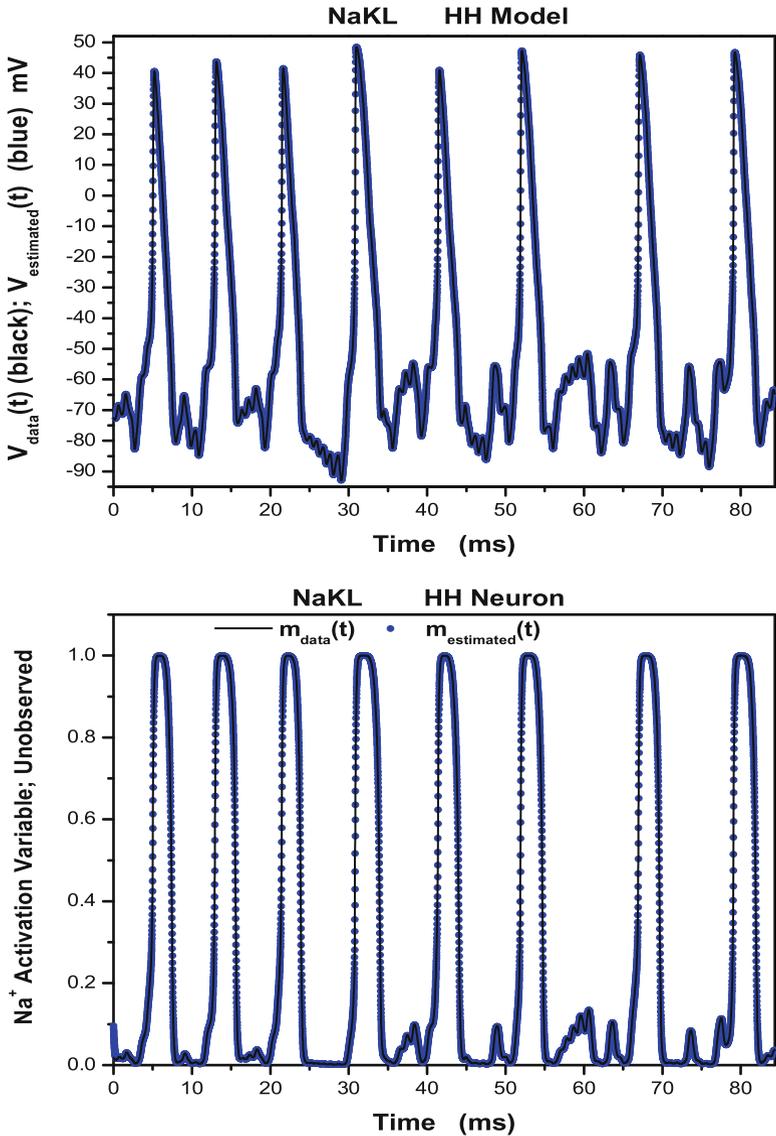


Fig. 2.16 NaKL model. *Top* We display the known and estimated membrane voltage $V_{\text{data}}(t), V_{\text{estimated}}(t)$ from “data” generated in a twin experiment using the NaKL model as a data source. *Bottom* We display the known and estimated Na^+ activation gating variable $m_{\text{data}}(t), m_{\text{estimated}}(t)$ from “data” generated in a twin experiment using the NaKL model as a data source. In an experiment $m(t)$ would be an unobserved state variable. Here the unobserved variables are known to us and comparison is possible

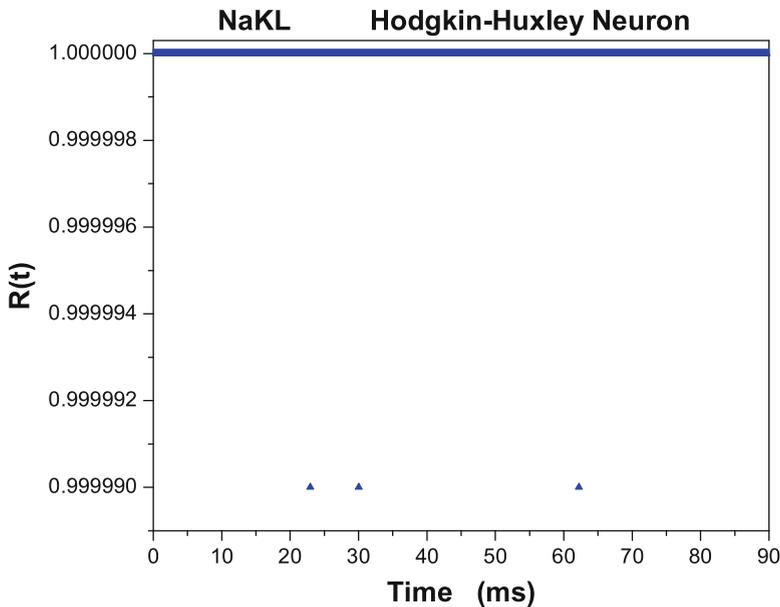


Fig. 2.17 $R(t)$, Eq. (2.45), over the observation window after the estimation procedure has completed. When $R(t)$ is near unity, the data and the model are consistent, and the role of $u(t)$ has been to regularize the search

encounters a different forcing function $I_{\text{app}}(t)$. This is not universally agreed on Wunsch and Heimbach (2007) using the methods described in this book or other instantiations of the ideas here. It is argued that the goal is to establish that the state of the system is well approximated by the data assimilation procedure. However, as we will see as we proceed, that can be quite misleading.

Our metric for success in the data assimilation will be the ability of the completed model to predict future response to different forcing. In order to accomplish this we require all the parameters \mathbf{p} in the model, thus completing it, and all of the state variables at the end of the observation window so the dynamical equations, with $u(t) = 0$, may be used to predict forward in time. This gives a rationale to the minimization algorithms we use, and which we will discuss in another section of the book. These estimate the state at each point in the observation window, including the endpoint $t = T$.

We have taken from our numerical optimization the estimation of the four quantities $\{V(T), m(T), h(T), n(T)\}$ and the many fixed parameters and used the dynamical equations to predict forward for $t > T = 90$ ms. In Fig. 2.18 we show in the Top Panel a repeat of the known voltage (black) and the estimated voltage (red) in the observation window, and in the Bottom Panel, we display the known voltage in black and the predicted voltage in blue. This is a twin experiment, $R(t) \approx 1$, and $V(t)$ is an observed variable appearing explicitly in the cost function, so we may not be surprised at this juncture at the accuracy of this result.

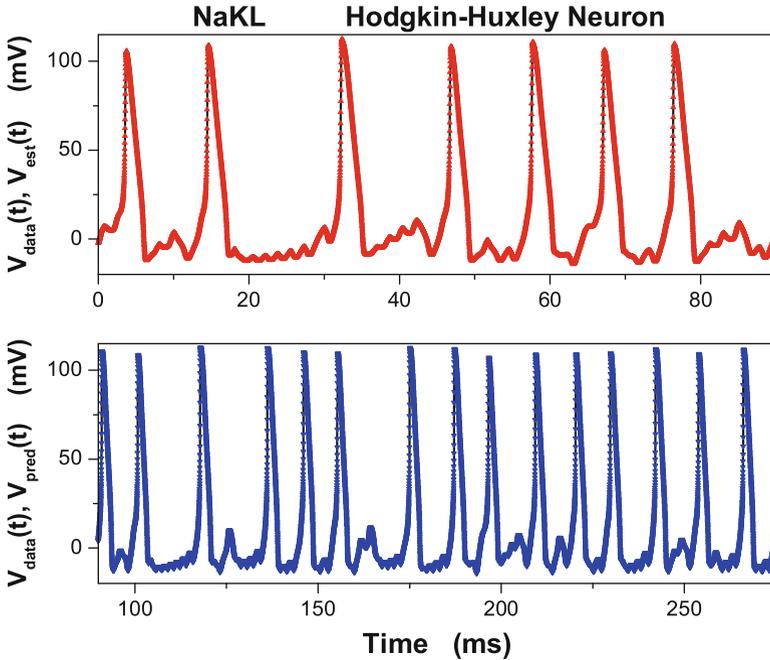


Fig. 2.18 NaKL model. *Top Panel* The “data” voltage was presented to the NaKL model. This shows those “data” and the estimated output voltage from the model. *Bottom Panel* This displays “data” voltage presented to the NaKL model and the predicted voltage after the observations are completed. This is an observed state variable

To further inquire into the method, we display in Fig. 2.19 the same type of result for an **unobserved** state variable. Here we show the K^+ gating variable $n(t)$. In black is the known data generated using the model equations with known parameters. Also displayed is the estimation (in red) of $n(t)$ in the observation window $[0, 90\text{ ms}]$ and the prediction for $t > 90\text{ ms}$ in blue. The accuracy of the results is consistent with the $R(t)$ test and speaks for itself.

2.3.4 Consequences of the Wrong Model

We address one last question of an introductory nature. If our procedure is so clever, can it give us signals of the model being wrong? To make this inquiry, we presented our NaKL model with membrane voltage “data” from a different neuron model which we call the NaKLh model. The biophysical difference was that the “wrong” data came from a neuron model with an additional voltage-gated current active primarily in the subthreshold voltage region. Using this wrong data, we carried out the same estimation procedure as above using 76.3 ms of “data”.

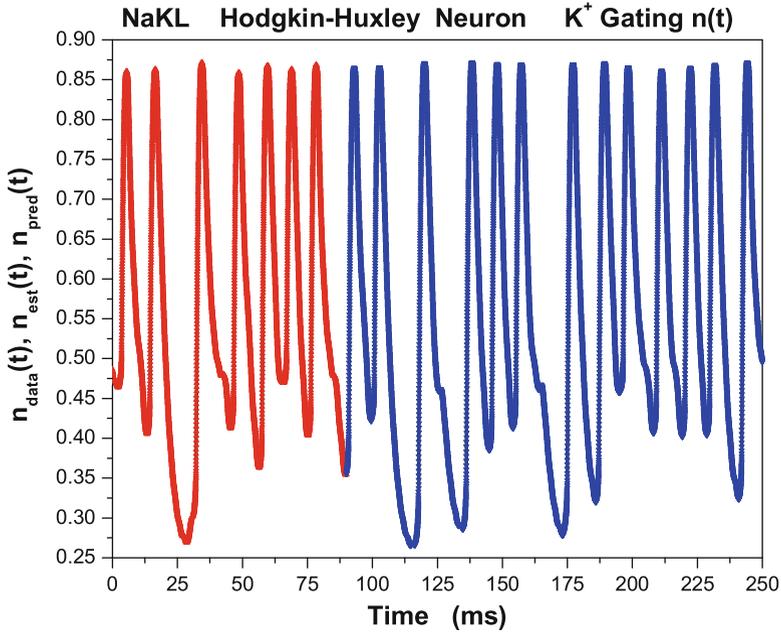


Fig. 2.19 NaKL model. Estimating and predicting the K^+ gating variable $n(t)$. The “data” $n_{\text{data}}(t)$ are in black; the estimated values $n_{\text{est}}(t)$ are in red; the predicted values $n_{\text{pred}}(t)$ using the 18 estimated parameters and the values of $\{V(T), m(T), h(T), n(T)\}$ at $T = 90$ ms and the dynamical equations with $u(t) = 0$. This is an unobserved state variable

In Fig. 2.20 we see the estimated voltage (blue) as well as the known voltage (black) presented to the model. The current stimulating the neuron model is essentially the same as before. While this estimation appears rather accurate, we take away a quite different conclusion when we examine the values of $R(t)$ in the observation window. These are displayed in Fig. 2.21 in which we observe significant deviations from $R(t) \approx 1$. In this twin experiment, further evidence that something is awry comes from examining the estimation of the Na^+ gating variable $m(t)$ in the observation window (Fig. 2.22). The significant disagreement between the known $m(t)$ and the estimated $m(t)$ is associated with the fact that the nonlinear filter of voltage information, namely, the model neuron, is incorrect. Of course, this indicator is not available for experimental data. Finally, though we do not display it, the predictions of the model are also incorrect in this situation.

2.4 Synopsis and Perspectives: “Slightly Complex” Examples

The examination of three low-dimensional example problems from fluid dynamics, electronic circuits and neurobiology has illustrated many of the issues addressed in this monograph. We have been presented in each case with one time series of

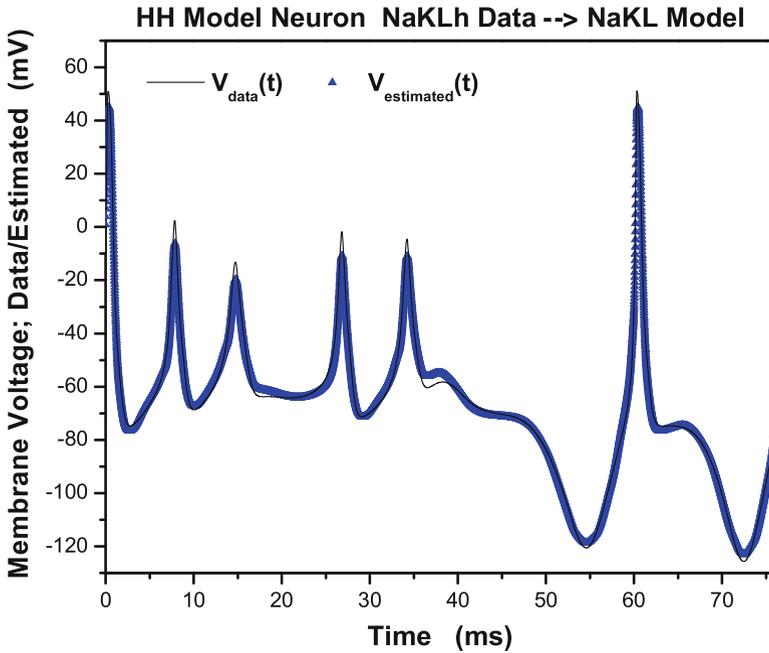


Fig. 2.20 The voltage data, *black line*, and the estimated voltage, *blue triangles*, when we present the “wrong” voltage data to an NaKL neuron model. There is an additional current in the data voltage and it is not represented in the simpler NaKL model. Nonetheless, the estimation procedure appears to produce a good representation of the presented data

measured data $y(t)$ from a physical system, and recognizing that the state of the system is represented by a model with a D -dimensional state variable $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_D(t)]$ $D \geq 1$, we sought to determine the fixed parameters \mathbf{p} of the model as well as the unobserved states of the model at some time T when observations were completed. Using the information in the data we showed in each case that selecting a metric to compare the data $y(t)$ with a physically motivated function of the state $h(\mathbf{x}(t))$, we could try to minimize the distance associated with the metric, constrained by the dynamical equations of the model, and successfully determine \mathbf{p} and the unobserved states. We also demonstrated that for $t > T$, we could predict the state of the model system.

When one has a measurement function $h_l(\mathbf{x})$ expressing the relation between the observed quantity and the model state, the cost function contains terms of the form $[y_l(t) - h_l(\mathbf{x}(t))]$, and in the use of variational principles as we used in this chapter, one requires the derivatives of the measurement function itself. This can be an issue when the measurement function has thresholds or very sharp variations. However, when we arrive at the discussion of the direct evaluation of the path integral representing the full answer to statistical data assimilation questions, these derivatives of measurement functions are avoided, but more about that as we proceed.

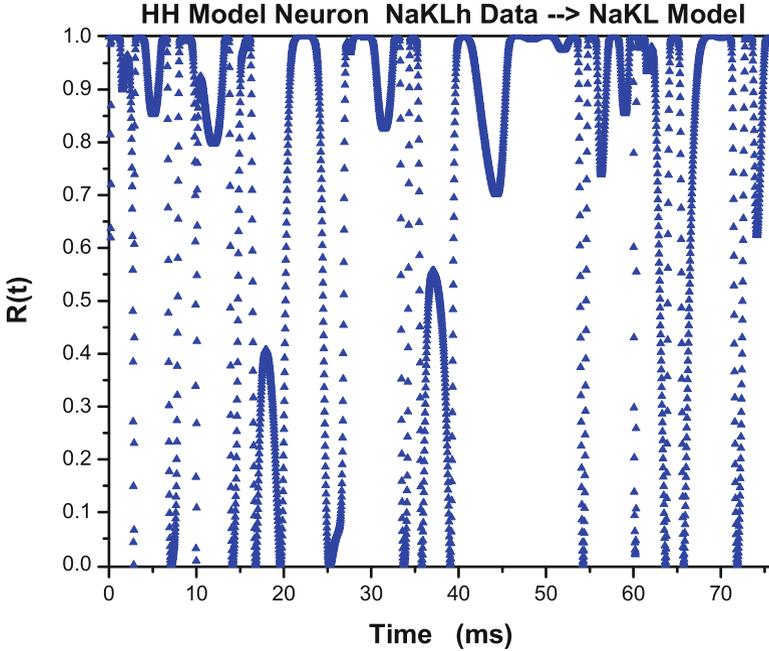


Fig. 2.21 The dimensionless ratio $R(t)$ resulting when an NaKL model is presented with data from a neuron model with an additional current, the NaKLh model. This shows that the NaKL model is inconsistent with the data

In our discussion in this chapter we made the simplifying assumption that $h(\mathbf{x}(t)) = x_1(t)$ and that the metric to use is a least-squares distance between the data and the corresponding model output

$$\begin{aligned}
 C(\mathbf{p}) &= \frac{1}{T} \int_0^T dt (y(t) - x_1(t))^2 \\
 &= \frac{1}{m+1} \sum_{n=0}^m (y(t_n) - x_1(t_n))^2.
 \end{aligned} \tag{2.46}$$

In the second form we recognized that measurements were made at the $m+1$ discrete times $t_n = n\Delta t$; $T = m\Delta t$. In addition to this standard formulation of the problem of estimations of parameters and state variables, we encountered a difficulty when the data time series and the model output were chaotic, and a regularization procedure was suggested that smoothed out the irregular plethora of local minima in $C(\mathbf{p})$ during the estimation period and then disposed of the regulation when the estimation was completed.

A number of questions arise about the procedures we have employed in this introductory chapter:

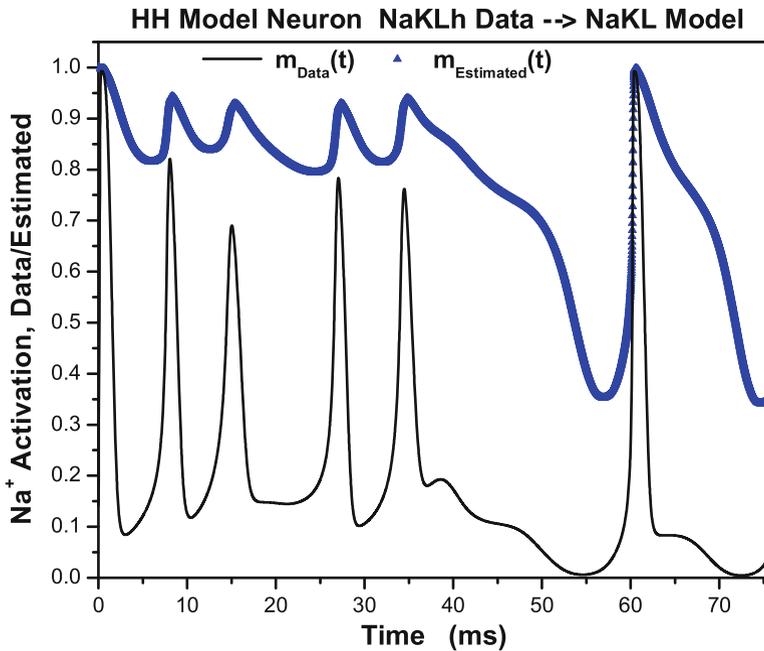


Fig. 2.22 The known Na^+ gating variable $m(t)$, black line, and the estimated $m(t)$, blue triangles, when data from the “wrong” model (NaKLh) is presented to an NaKL model. Since this is a twin experiment, we are able to compare the actual $m(t)$ in the data with the $m(t)$ estimated by the variational principle. This type of comparison is absent in a real experiment where only membrane voltage is observed

- We assumed that a least-squares metric to compare our data $y(t)$ to the model output $x_1(t)$ should transmit enough information to enable estimation of the parameters and unobserved states in the models. What formulation replaces that when the statistics of the data are not Gaussian?
- We more or less assumed that passing measurements from one of the system state variables to the model would be enough to enable accurate estimation of the model parameters and the model’s unobserved states. When does that remain true for more complex systems? When are more observations at each measurement time required?
- We estimated unobserved model states and model parameters using an optimization principle with the model dynamical equations as equality constraints. What is the correct rule when the model, as well as the observations, has errors?
- We more or less assumed that when we estimated the state values, in particular the initial conditions $\mathbf{x}(t_0)$ when observations begin, those could be determined precisely. Is this always so? What role does uncertainty in the initial conditions play in the estimation task?

- We regularized the instabilities on the synchronization manifold with the addition of a control $u(t)$ which drives the model output to the data, then is itself driven to zero by the optimization procedure. This appears ad hoc. What other approaches can one use?
- We took **prediction** for times beyond the observation or assimilation window $t > T$ as the metric for a successful analysis of a model provided with data from observations. This goes beyond a “good fit” of model output to observed data as it requires the full state of the model to be accurately estimated at the completion of observations, and it uses the nonlinear model as a filter of information in the data to be utilized in determining the unobserved state variables. Are there better metrics for model completion and correctness? Can one develop criteria for testing the consistency of the model with the available data?

We do not have complete answers to these issues and many others that arise as we proceed. Nonetheless, we now turn to the general formulation of data assimilation questions and methods for answering them in a practical fashion.

Chapter 3

General Formulation of Statistical Data Assimilation

Prologue to the Chapter

In this chapter we give the general formulation of the statistical problem of transferring information from observations to a model of the observed system. This is named data assimilation, and when the measurements are noisy, the models have errors, and we know only a distribution of possible states of the model when observations are initiated, the problem is statistical. The discussion begins by assuming we have no observations, and this is the arena for us to discuss how we propagate a model with errors forward in time. This is the classical analog of the more familiar quantum mechanical problem and, as with the quantum question, is answered via an integral representation for the statistical description of the model state within some interval $[0, T]$. This is a path integral in both quantum and classical considerations. In quantum theory it is the probability amplitude we seek, while in classical statistical physics it is a real probability distribution we want.

We then permit measurements in the time interval $[0, T]$ and using identities on conditional probabilities uncover the precise rule entering the path integral associated with a measurement at $t_n; 0 \leq t_n \leq T$. We illustrate this in some detail with one, then two, then many measurements. Within this discussion we recognize the importance of making quantitative the manner in which measurements are made. This requires a statement of the relation between the state variables of the model and the quantities reported in the observations.

The results of this chapter provide the general framework for answering all questions associated with statistical data assimilation. Further chapters report on various methods to evaluate the integral representation uncovered here.

3.1 Data Assimilation Without Data

3.1.1 Deterministic Dynamics: Path Integral

We turn now to general questions about transferring information from measurements to a quantitative model of the processes producing those data. The setting we envision is that of experiments in a laboratory or in the field where the observations are noisy, where our quantitative models have errors, and where the state of the model system is uncertain when we begin observations. This means that all questions are now statistical, and we must address probability distributions for states and parameters. Observables are then the various expectation values of moments of the states and variations about these moments.

Let's begin without any acquisition of data. To proceed we need to choose a state space for the description of the data through a quantitative model. If the data describe the deterministic dynamics as a function of space as well as time, the physical description would be of a field $\Phi_\alpha(\mathbf{r}, t)$; $\mathbf{r} = \{r_1, r_2, r_3\}$; $\alpha = 1, 2, \dots$. Satisfying partial differential equations (PDEs)

$$\frac{\partial \Phi_\alpha(\mathbf{r}, t)}{\partial t} = \mathcal{F}_\alpha(\Phi(\mathbf{r}, t)), \quad (3.1)$$

α is a vector index enumerating the components of the field.

Analysis of these equations may proceed by first discretizing space

$$\Phi_\alpha(\mathbf{r}, t) \rightarrow \Phi_\alpha(i \Delta r_1, j \Delta r_2, k \Delta r_3, t). \quad (3.2)$$

$\{i, j, k\}$ are integers, and $\{\Delta r_1, \Delta r_2, \Delta r_3\}$ are grid increments in three-dimensional space. This results in ordinary differential equations for $x_a(t) = \Phi_\alpha(i \Delta r_1, j \Delta r_2, k \Delta r_3, t)$; a is the collection of labels $\{i, j, k, \alpha\}$:

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t), \mathbf{p}), \quad a = 1, 2, \dots, D \quad (3.3)$$

with \mathbf{p} a vector of time-independent parameters. If we also discretize time $t \rightarrow \{t_0, t_1, \dots, t_n, \dots\}$, this gives us dynamics in the form of relations between the state at time t_n and the state at time t_{n+1} :

$$g_a(\mathbf{x}(t_{n+1}), \mathbf{x}(t_n), \mathbf{p}) = 0. \quad (3.4)$$

The time increments $t_{n+1} - t_n$ need not all be identical. In a shorthand where $\mathbf{x}(t_n) = \mathbf{x}(n)$, the deterministic dynamics is $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$. If the dynamics can be written in explicit form, $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \rightarrow x_a(n+1) - f_a(\mathbf{x}(n), \mathbf{p})$. The function $\mathbf{f}(\mathbf{x}(n), \mathbf{p})$ is often the algorithm for solving the differential equations (3.3) in discrete time.

Now we recognize that our knowledge of the state of the model at t_0 could also be uncertain, requiring us to introduce the probability distribution for $\mathbf{x}(t_0) : P(\mathbf{x}(0))$. We wish to follow that distribution of initial states through a window of time $\{t_0, t_1, \dots, t_n, \dots, t_m = T\}$, and for this purpose we want to propagate the probability distribution function $P(\mathbf{x}(t))$ of the model with a state $x_a(t_n) = x_a(n)$; $a = 1, 2, \dots, D$ from time t_0 to time t_1 to ... and finally to time $t_m = T$. At the final time t_m we want to know the distribution of states $P(\mathbf{x}(t_m)) = P(\mathbf{x}(m))$.

To move along this path in time, the basic quantity to address is the joint probability distribution in $D(m+1)$ -dimensional space $P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0))$ of the states $\mathbf{x}(t_n)$ at each discrete time t_n in the interval. We are asking about the outcome of an initial value problem where, given, $P(\mathbf{x}(0))$, we seek to determine $P(\mathbf{x}(m))$. In continuous time, the equation governing this time development is a PDE for the probability density involving only linear terms in the time derivative (Lindenberg and West 1990), so $P(\mathbf{x}(0))$ suffices, along with a statement of the dynamics, to give $P(\mathbf{x}(t_n)) = P(\mathbf{x}(n))$ for $n > 0$.

We now assume that the temporal progress of the system is such that the state of the system $\mathbf{x}(n+1)$ at time t_{n+1} is determined completely by the state of the system at time t_n , $\mathbf{x}(n)$. This is the situation described by the usual partial (or ordinary) differential equations representing physical and biophysical processes and inherited through the steps discretizing space and time. Dynamical systems where $\mathbf{x}(n+1)$ is established by $\mathbf{x}(n)$ are known as Markov, and as they naturally arise when differential equations describe dynamics, they are quite well studied.

When would we **not** encounter this Markov property in the discussion of physical or biological processes? This might happen when we have started with a large space of states and projected it down to D -dimensions in one way or another. One example would be an active optical medium connected to itself via a length l of optical fiber. The propagation of light from the active medium to itself via the fiber is that of an electromagnetic wave moving along a glass waveguide. If it propagates without substantial interaction with the fiber environment, we might replace the PDEs for wave propagation in the fiber by examining the active medium as a device emitting a signal into the fiber at some time t and receiving that signal back as input at a time $t+l/c$. This replaces the fiber details by a time delay device, and now the dynamical equations of the active medium alone connect its state $\mathbf{x}(t)$ with the output of the its earlier state $\mathbf{x}(t-l/c)$. Such delay differential equations are formally infinite dimensional as one requires all values of \mathbf{x} between some initial time t_0 and the time t_0-l/c to integrate them forward in time. They are not Markov, but, of course, we imposed that on ourselves by projecting out the infinite-dimensional PDEs (Maxwell's equations in this example) moving the light through the fiber.

We would also not have a Markov process if we encountered a situation where the state $\mathbf{x}(n+1)$ depended on $\mathbf{x}(n)$, as would be usual for a Markov process, and also depended on $\mathbf{x}(n-1)$ and possibly the states at yet earlier times. By enlarging the state space to include the finite number of such delays, we would again produce Markov dynamics in the larger space. In the fiber cable example, there are an infinite number of delays, but the setting becomes Markov again when we restore the PDEs for moving light about.

With these caveats, we proceed assuming that our model has the Markov property. This means that the probability distribution in $D(m+1)$ -dimensional space may be written as

$$\begin{aligned}
& P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)) \\
&= \frac{P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0))}{P(\mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0))} P(\mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)) \\
&= P(\mathbf{x}(m)|\mathbf{x}(m-1), \mathbf{x}(m-2), \dots, \mathbf{x}(0)) P(\mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)) \\
&= P(\mathbf{x}(m)|\mathbf{x}(m-1)) P(\mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)). \tag{3.5}
\end{aligned}$$

We have used the common notation for a conditional probability: if the distribution of a quantity q is conditional on a quantity ξ , $P(q|\xi)$, then one denotes this as $P(q|\xi) = P(q, \xi)/P(\xi)$, where $P(q, \xi)$ is the joint probability distribution for q and ξ .

The distribution of the state $\mathbf{x}(m)$ in the conditional probability $P(\mathbf{x}(m)|\mathbf{x}(m-1), \mathbf{x}(m-2), \dots, \mathbf{x}(0))$, $\mathbf{x}(m)$ is determined only by $\mathbf{x}(m-1)$ and is independent of $\mathbf{x}(m-2), \mathbf{x}(m-3), \dots$ allowing us to write

$$P(\mathbf{x}(m)|\mathbf{x}(m-1), \mathbf{x}(m-2), \dots, \mathbf{x}(0)) = P(\mathbf{x}(m)|\mathbf{x}(m-1)) \tag{3.6}$$

in the third line of Eq. (3.5). This is the explicit consequence of the Markov property. Continuing along the chain of Markov processes, we arrive at

$$\begin{aligned}
& P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)) \\
&= P(\mathbf{x}(m)|\mathbf{x}(m-1))P(\mathbf{x}(m-1)|\mathbf{x}(m-2))P(\mathbf{x}(m-2), \mathbf{x}(m-3), \dots, \mathbf{x}(0)) \\
&= \prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)) P(\mathbf{x}(0)). \tag{3.7}
\end{aligned}$$

The quantity $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ is called the transition probability. It is a $D \times D$ matrix in the space with time labels fixed, and it tells us how we move along in time from state $\mathbf{x}(n)$ to $\mathbf{x}(n+1)$. The transition probability is the dynamics $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ stated in a statistical setting.

The probability density for the state at the end of the Markov chain $P(\mathbf{x}(m))$ results from integrating the joint distribution $P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0))$ of the $\mathbf{x}(n)$'s over all times before m

$$\begin{aligned}
P(\mathbf{x}(m)) &= \int \prod_{n=0}^{m-1} d^D x(n) P(\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(n), \dots, \mathbf{x}(0)) \\
&= \int \prod_{n=0}^{m-1} d^D x(n) P(\mathbf{x}(n+1)|\mathbf{x}(n)) P(\mathbf{x}(0)). \tag{3.8}
\end{aligned}$$

This has the interpretation that the probability distribution for the final state $P(\mathbf{x}(m))$ is reached by taking the distribution at t_0 , $P(\mathbf{x}(0))$ and first propagating it forward one step to t_1 , yielding $P(\mathbf{x}(1))$:

$$P(\mathbf{x}(1)) = \int d^D x(0) P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)). \quad (3.9)$$

We follow this action by moving along to time t_2 , where we encounter

$$P(\mathbf{x}(2)) = \int d^D x(0) d^D x(1) P(\mathbf{x}(2)|\mathbf{x}(1)) P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)), \quad (3.10)$$

and then proceed in the same manner until we reach time t_m . The probability distribution at the end of this chain of Markov processes, $P(\mathbf{x}(m))$, is now seen as an integral along a path $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(m)\}$ in the D -dimensional space of the state of the system as it moves from time t_0 to time t_1, \dots , until it reaches the end of the path at time t_m . At each time step we propagate the state \mathbf{x} forward by the dynamics embodied in the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$.

For comparison to other well-studied path integral problems (Feynman and Hibbs 1965; Zinn-Justin 2002), we write this path integral, Eq. (3.8), as

$$\begin{aligned} P(\mathbf{x}(m)) &= \int \prod_{n=0}^{m-1} d^D x(n) P(\mathbf{x}(n+1)|\mathbf{x}(n)) P(\mathbf{x}(0)) \\ &= \int \prod_{n=0}^{m-1} d^D x(n) \exp[-A_0(\mathbf{X})], \\ &= \int d^D x(0) K(\mathbf{x}(m), \mathbf{x}(0)) P(\mathbf{x}(0)), \end{aligned} \quad (3.11)$$

with $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(m)\}$, a vector in $(m+1)D$ -dimensional space. $A_0(\mathbf{X})$ is

$$A_0(\mathbf{X}) = - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))], \quad (3.12)$$

and the “kernel” is defined by

$$\begin{aligned} K(\mathbf{x}(m), \mathbf{x}(0)) &= \int \prod_{n=1}^{m-1} d^D x(n) \prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)) \\ &= \int d\mathbf{X} \prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)). \end{aligned} \quad (3.13)$$

We refer to $A_0(\mathbf{X})$ as the ‘‘action’’. The analogy with the classical action in the quantum mechanical counterpart suggests this name.

Any function along the path $\mathbf{X}\{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(m)\}$, $G(\mathbf{X})$, has the expected value

$$\begin{aligned} \langle G(\mathbf{X}) \rangle &= \int d\mathbf{X} G(\mathbf{X}) \prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)) \\ &= \frac{\int d\mathbf{X} \exp[-A_0(\mathbf{X})] G(\mathbf{X})}{\int d\mathbf{X} \exp[-A_0(\mathbf{X})]}. \end{aligned} \quad (3.14)$$

3.1.2 Relation to the Quantum Mechanical Path Integral

The argument for this kind of path integral is found most often in discussions of quantum dynamics where the probability distribution $P(\mathbf{x}(n))$ is replaced by a complex probability amplitude $\psi(\mathbf{x}, t_n)$ and the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ is replaced by the quantity

$$P(\mathbf{x}(n+1)|\mathbf{x}(n)) \longrightarrow K(\mathbf{x}(n+1), \mathbf{x}(n)) = e^{\frac{i}{\hbar} S[t_{n+1}, t_n]},$$

where

$$S[t_{n+1}, t_n] = \int_{t_n}^{t_{n+1}} dt L(\mathbf{q}(t), \dot{\mathbf{q}}(t), t) \quad (3.15)$$

and $L(\mathbf{q}(t), \dot{\mathbf{q}}(t), t)$ is the classical Lagrangian for the motion of the object with coordinates $\{\mathbf{q}(t), \dot{\mathbf{q}}(t)\}$ (Feynman and Hibbs 1965; Zinn-Justin 2002). In the quantum mechanical case one has

$$\psi(\mathbf{x}(m), t_m) = \int \prod_{n=0}^{m-1} d^D x(n) K(\mathbf{x}(n+1), \mathbf{x}(n)) \psi(\mathbf{x}(0), t_0). \quad (3.16)$$

The direct analogy between the propagation of the real probability distribution of the classical state \mathbf{x} along a path moved along by the nonlinear dynamics $\mathbf{x}(n) \rightarrow \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ and the evolution of a complex wave function arises because each satisfies a linear PDE. For the wave function that is Schrödinger’s equation and for the classical probability distribution, it is the master equation (Lindenberg and West 1990).

3.1.3 Noisy Dynamics

The probability distribution in noise-free, deterministic movement along the chain of Markov processes is determined by the transition probability

$$\begin{aligned} P(\mathbf{x}(n+1)|\mathbf{x}(n)) &= \delta^D(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})) \\ P(\mathbf{x}(n+1)|\mathbf{x}(n)) &= \delta^D(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p})). \end{aligned} \quad (3.17)$$

The sharp delta function is broadened in the case of uncertain dynamics. There is not a unique prescription for the broadening of the delta function. If we regard the source of this broadening as loss of perfect resolution in the deterministic dynamics, we probably should use one of the many representations of a delta function, each of which as a width associated with the loss of resolution. However, if the replacement of the delta function in representing the transition probability is due to terms simply missing from our formulation of those dynamics, no easy rule is known.

Adding noise $\theta(n)$ to the dynamics at each time t_n changes the deterministic expression to

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p}) + \theta(n). \quad (3.18)$$

The transition probability for a given realization of θ is $\delta^D(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p}) - \theta(n))$. When we average any function of $\mathbf{X} = \{x_a(0), x_a(1), \dots, x_a(m)\}$, which now depends on θ through Eq.(3.18), over the distribution of θ , $P_\Theta(\theta)$, the expectation value of a function on the path $G(\mathbf{X})$ is expressed as

$$\begin{aligned} \langle G(\mathbf{X}) \rangle_\theta &= \int \prod_{n=0}^m d\theta(n) P_\Theta(\theta(n)) \\ &\quad \prod_{n=0}^{m-1} d^D x(n) P(\mathbf{x}(n+1)|\mathbf{x}(n)) P(\mathbf{x}(0)) G(\mathbf{X}) \\ &= \int \prod_{n=0}^m d\theta(n) P_\Theta(\theta(n)) \\ &\quad \prod_{n=0}^{m-1} d^D x(n) \delta^D(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p}) - \theta(n)) P(\mathbf{x}(0)) G(\mathbf{X}) \\ &= \prod_{n=0}^{m-1} d^D x(n) P_\Theta(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p})) P(\mathbf{x}(0)) G(\mathbf{X}). \end{aligned} \quad (3.19)$$

It has the same form as Eq. (3.14) but with

$$A_0(\mathbf{X}) = - \sum_{n=0}^{m-1} \log \left[P_\Theta(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p})) \right] - \log[P(\mathbf{x}(0))]. \quad (3.20)$$

If the noise term is state dependent, one may change to dynamical variables in which it becomes additive (West et al. 1979).

Path integrals without data acquisition (Feynman and Hibbs 1965; Zinn-Justin 2002) have an enormous literature. The path integral is an integral representation of the solution to a PDE for $P(\mathbf{x}(t))$, and as with other integral representations of interesting quantities, it permits a global view of that probability distribution (or wave function) and allows discussions of the transit along a whole path of states $\mathbf{x}(t)$. All the properties contained in the integral representation of $P(\mathbf{x}(t))$ are found in the original differential equation, though the local view in space may not reveal the same informative ways to approximate the solution.

3.2 Data Assimilation with a Little Bit of Data

The propagation of a probability distribution forward in time brings us part of the way toward our goal. We want to formulate the same activity when observed data are available to provide information to the distribution (or probability amplitude) at a selected set of times within the interval $[t_0, t_m = T]$. To identify how the Markov rule for chains of events is altered through the presence of information transfer, we begin with a consideration of a single piece of data $\mathbf{y}(t)$. This is a vector of L measurements at the time t : $\mathbf{y}(t) = \{y_1(t), y_2(t), \dots, y_L(t)\}$. In this consideration, we will encounter the notion of conditional mutual information (Fano 1961), and we now take a step sideways to consider that.

3.2.1 Mutual Information

As explained in pedagogically useful detail in the book by Fano (1961), a measure of information about the occurrence of an event a_k which is an element of a set of quantities $A = \{a_k\}$ and its connection with the occurrence of an event b_n which is an element of a set of quantities $B = \{b_n\}$ is given in terms of the joint probability distribution of events a_k and b_n , $P_{AB}(a_k, b_n)$, and the marginal distributions $P_A(a_k) = \sum_{b_n} P_{AB}(a_k, b_n)$ and $P_B(b_n) = \sum_{a_k} P_{AB}(a_k, b_n)$ via the mutual information $\text{MI}(a_k, b_n)$:

$$\begin{aligned} \text{MI}(a_k, b_n) &= \log \left\{ \frac{P_{AB}(a_k, b_n)}{P_A(a_k) P_B(b_n)} \right\} = \log \left\{ \frac{P_{A|B}(a_k|b_n)}{P_A(a_k)} \right\} = \log \left\{ \frac{P_{B|A}(b_n|a_k)}{P_B(b_n)} \right\} \\ &= \text{MI}(b_n, a_k). \end{aligned} \tag{3.21}$$

This quantity answers the question: how much (in bits with logarithms to base 2) do we learn about the event b_n if we observed the event a_k ? This is symmetric in the a_k and b_n .

The MI varies over the elements of the sets A and B , so it is a statistic over these sets. The average over the elements of these sets is the average mutual information:

$$\text{AMI}(A, B) = \sum_{\{a_k, b_n\}} P_{AB}(a_k, b_n) \text{MI}(a_k, b_n) \geq 0. \quad (3.22)$$

$\text{AMI}(A, B)$ plays a central role in the discussion of the capacity of a statistical communications channel (Fano 1961). It is often, in a relaxed fashion, itself called the mutual information. However, when the distributions entering its definition are multimodal or broad, averages or higher-order moments of $\text{MI}(a_k, b_n)$ may not characterize the statistical elements $\text{MI}(a_k, b_n)$ very well.

Indeed, $\text{MI}(a_k, b_n)$ is not always positive, though its average $\text{AMI}(A, B)$ over the sets A and B is positive or zero (Fano 1961). $\text{MI}(a_k, b_n)$ is zero only when the sets A and B are independent and $P_{AB}(a_k, b_n) = P_A(a_k) P_B(b_n)$.

When there are more than two relevant sets of variables, say $A = \{a_k\}$, $B = \{b_n\}$ and $C = \{c_j\}$, then the conditional mutual information among conditional probability distributions is given as

$$\text{CMI}(a_k, b_n | c_j) = \log \left\{ \frac{P_{AB|C}(a_k, b_n | c_j)}{P_{A|C}(a_k | c_j) P_{B|C}(b_n | c_j)} \right\}. \quad (3.23)$$

This is the information about the observation of a_k and b_n conditioned on having observed c_j . It answers the question: how much is learned about the observation b_n upon making the observation a_k conditioned on having observed c_j ? If the logarithm is taken to base 2, the answer is in bits. It is precisely the quantity we will encounter in assessing the transfer of information from a data source to a dynamical model of the source.

Both the MI and the CMI are functions varying over the sets of variables $\{a_k\}$, $\{b_n\}$, $\{c_j\}$, \dots , and they are well defined when the prescription that $x \log(x) \rightarrow 0$ when $x \rightarrow 0$ is adopted. In most books, the variables are referred to as random, but, in fact, the set of variables need only be a collection of varying quantities. If the elements a_k, \dots are drawn from locations on the attractor of a dissipative dynamical system, they are deterministic, and that is not random at all. Nonetheless, these information theoretic quantities are well defined and meaningful.

Mutual Information: A Symmetric Binary Channel Example

To illustrate some of these points, we use an example taken from Fano (1961). In Fig. 3.1 we show a communications channel from a source that can take two values $\{s_0, s_1\}$ with probabilities p_s and $1 - p_s$. This means the source is sending signals through the channel, and the signal s_0 occurs with a frequency p_s , while the signal s_1 occurs with a frequency $1 - p_s$. There is a transmission probability of the channel called t , and this is the frequency with which a signal emitted $\{s_0, s_1\}$ from the source arrives at the receiver as $\{r_1, r_0\}$. The frequency with which the source signal $\{s_0, s_1\}$ arrives at the receiver as $\{r_0, r_1\}$ is $1 - t$.

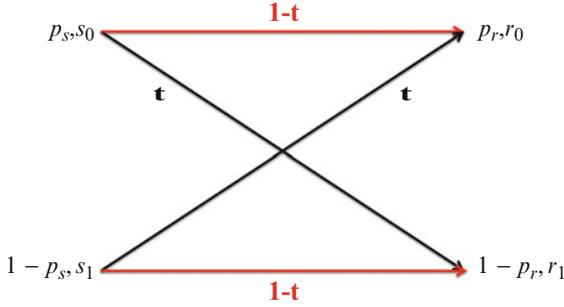


Fig. 3.1 Illustration of the binary symmetric information channel. There are two source states $\{s_0, s_1\}$ occurring with probabilities p_s and $1-p_s$, respectively, and there are two response channels $\{r_0, r_1\}$ occurring with probabilities p_r and $1-p_r$, respectively. The probability of the transition $s_0 \rightarrow r_1$ during the transmission of information from source to response is t , as is the probability of $s_1 \rightarrow r_0$. $s_0 \rightarrow r_0$ and $s_1 \rightarrow r_1$ each occur with probability $1-t$

From the figure we conclude that the frequency with which r_0 is seen at the receiver given that s_0 departed the source is $P(r_0|s_0) = 1-t$. The frequency with which r_0 is seen at the receiver given that s_1 departed the source is $P(r_0|s_1) = t$. The frequency with which r_1 is seen at the receiver given that s_0 departed the source is $P(r_1|s_0) = t$. The frequency with which r_1 is seen at the receiver given that s_1 departed the source is $P(r_1|s_1) = 1-t$. From this and the definition of conditional probability, we have

$$\begin{aligned}
 P(s_0, r_0) &= P(r_0|s_0)P(s_0) = p_s(1-t) & P(s_0, r_1) &= P(r_1|s_0)P(s_0) = tp_s \\
 P(s_1, r_0) &= P(r_0|s_1)P(s_1) = t(1-p_s) & P(s_1, r_1) &= P(r_1|s_1)P(s_1) = (1-p_s)(1-t).
 \end{aligned} \tag{3.24}$$

From these results we can evaluate the mutual information $MI(s_a, r_b)$; $a, b = 1, 2$ for pairs of signals sent from the source and registered at the receiver:

$$\begin{aligned}
 MI(s_0, r_0) &= \log \left\{ \frac{1-t}{p_r} \right\} & MI(s_0, r_1) &= \log \left\{ \frac{t}{1-p_r} \right\} \\
 MI(s_1, r_0) &= \log \left\{ \frac{1}{p_r} \right\} & MI(s_1, r_1) &= \log \left\{ \frac{1-t}{1-p_r} \right\},
 \end{aligned} \tag{3.25}$$

and finally we determine the frequency p_r at which the signal r_0 is seen at the receiver: $p_r: p_r = p_s + t - 2p_s t = p_s(1-t) + t(1-p_s)$.

Expanding on this example, let us suppose the transmission frequency is $t = 1/5$ and the frequency with which we send s_0 from the source is $p_s = 4/7$. We have then $p_r = 19/35$ and

$P(s_0, r_0)$	$P(s_0, r_1)$	$P(s_1, r_0)$	$P(s_1, r_1)$
16/35	4/35	3/35	12/35
$MI(s_0, r_0)$	$MI(s_0, r_1)$	$MI(s_1, r_0)$	$MI(s_1, r_1)$
$\log_2[28/19]$	$\log_2[7/16]$	$\log_2[35/19]$	$\log_2[28/15]$

One of the mutual information elements is negative, $\text{MI}(s_0, r_1) = -1.193$; nonetheless, the average mutual information (Fano 1961)

$$\text{AMI}(S, R) = \sum_{\{s=s_0, s_1\}, \{r=r_0, r_1\}} P(s, r) \text{MI}(s, r) \quad (3.26)$$

is positive.

This completes our excursion into the idea of conditional mutual information. There are many books on information theory that provide further discussion and often applications to the theory of communications.

3.2.2 One Measurement

So far we have considered propagation in time of Markov processes with no measurements. Next we examine what happens to the conditional probability distribution of the state of the model when we make one measurement. We will require some identities on conditional probabilities as well as the Markov property of the dynamics. We start with an example where we are concerned with the D -dimensional state at three times only: $\{\mathbf{x}(t_0), \mathbf{x}(t_1), \mathbf{x}(t_2)\} = \{\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2)\}$. An L -dimensional measurement is made at time t_1 with outcome $\mathbf{y}(1) = \{y_l(1)\}$; $l = 1, 2, \dots, L$. We seek the probability distribution at time t_2 conditioned on the measurement at t_1 . That is, we look for a representation of $P(\mathbf{x}(2)|\mathbf{y}(1))$.

We first note that by the Markov property of the dynamics taking us along the path $\mathbf{x}(0) \rightarrow \mathbf{x}(1) \rightarrow \mathbf{x}(2)$, we may go from t_1 to t_2 via

$$P(\mathbf{x}(2)|\mathbf{y}(1)) = \int d^D x(1) P(\mathbf{x}(2)|\mathbf{x}(1)) P(\mathbf{x}(1)|\mathbf{y}(1)), \quad (3.27)$$

as moving from time t_1 to time t_2 only requires the dynamics expressed in the form of the transition probability $P(\mathbf{x}(2)|\mathbf{x}(1))$. No measurements are made between those times. For clarity in placing measurements at a time location, we agree to make the measurement $\mathbf{y}(t_n)$ at an infinitesimal increment in time before t_n : namely, at $t_n - \epsilon$, letting $\epsilon \rightarrow 0$ after the calculation. Equation (3.27) is known as the Chapman–Kolmogorov equation (Chapman 1928; Kolmogorov 1931) and is true in general for Markov processes.

Using the properties of a conditional probability, we write

$$\begin{aligned} P(\mathbf{x}(1)|\mathbf{y}(1)) &= \frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} P(\mathbf{x}(1)) \\ &= \frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} \int d^D x(0) P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)), \end{aligned} \quad (3.28)$$

and putting this together with Eq. (3.27) we have for $P(\mathbf{x}(2)|\mathbf{y}(1))$

$$\begin{aligned} P(\mathbf{x}(2)|\mathbf{y}(1)) &= \int d^D x(1) d^D x(0) P(\mathbf{x}(2)|\mathbf{x}(1)) \\ &\quad \times \frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)) \\ &= \int d^D x(1) d^D x(0) P(\mathbf{x}(2)|\mathbf{x}(1)) \\ &\quad \times \exp[\text{MI}(\mathbf{x}(1), \mathbf{y}(1))] P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)). \end{aligned} \quad (3.29)$$

This tells us that one measurement made at time t_1 with outcome $\mathbf{y}(1)$ affects the movement down the chain of Markov processes by a factor which is precisely the exponential of the mutual information (Fano 1961)

$$\log \left\{ \frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} \right\} \quad (3.30)$$

between the L -dimensional measurement $\mathbf{y}(1)$ and the D -dimensional state $\mathbf{x}(1)$ at time t_1 . This result appears to have first been discussed by Lindley (1956).

As a result of the observations $\mathbf{y}(1)$ the action, Eq. (3.12), acquires an additional term

$$\begin{aligned} A_0(\mathbf{X}, \mathbf{y}(1)) &= -\text{MI}(\mathbf{x}(1), \mathbf{y}(1)) - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))], \\ A_0(\mathbf{X}, \mathbf{y}(1)) &= -\log \left[\frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} \right] - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))], \end{aligned}$$

in which the path through state space is $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2)\}$.

Any function on the path \mathbf{X} has a conditional expectation value

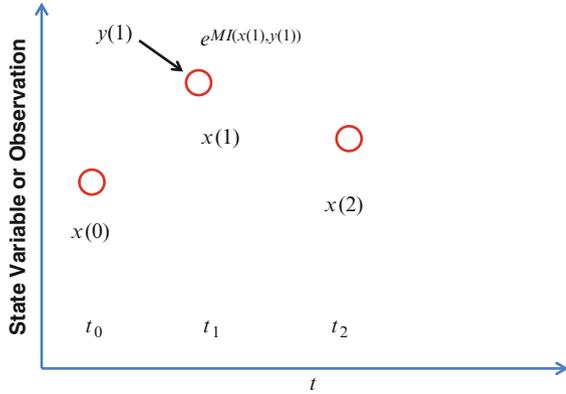
$$E[G(\mathbf{X})|\mathbf{y}(1)] = \frac{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{y}(1))] G(\mathbf{X})}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{y}(1))]}, \quad (3.31)$$

conditioned on the observation $\mathbf{y}(1)$ (Fig. 3.2).

The action can be slightly simplified in the evaluation of this ratio for any expectation value. We note that

$$\begin{aligned} A_0(\mathbf{X}, \mathbf{y}(1)) &= -\log \left[\frac{P(\mathbf{x}(1), \mathbf{y}(1))}{P(\mathbf{x}(1)) P(\mathbf{y}(1))} \right] \\ &\quad - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))] \end{aligned}$$

Fig. 3.2 Illustration of a measurement at time t_1 of the L quantities $\mathbf{y}(1)$ as the state of a model moves from $t_0 \rightarrow t_1 \rightarrow t_2$. The result of the measurement at t_1 is the introduction of a factor $\exp[\text{MI}(\mathbf{x}(1), \mathbf{y}(1))]$ into the Markov chain result associated with no measurements



$$\begin{aligned}
 &= -\log[P(\mathbf{y}(1)|\mathbf{x}(1))] - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))] \\
 &\quad + \log[P(\mathbf{y}(1))], \tag{3.32}
 \end{aligned}$$

and the last entry is independent of \mathbf{X} , so it cancels between the numerator and denominator of Eq. (3.31).

If we think of the unperturbed propagation of the model dynamics from time t_n to time t_{n+1} as given by $P(\mathbf{x}(n+1)|\mathbf{x}(n))$, then the term $\text{MI}(\mathbf{x}(1), \mathbf{y}(1))$ in the action acts as a guiding potential influencing the time development of the model system through the transfer of information from the observation $\mathbf{y}(1)$. The analogy suggests that the unguided movement of the state acts like the unperturbed dynamics of a nonlinear system, and the perturbation associated with a measurement guides the unperturbed motion through the transfer of mutual information to the dynamics through a subset of the full state at the time of a measurement. This information from the measurement moves the distribution of the state toward the state space location where $\mathbf{y}(1)$ is possible. The same feature is found in the quantum mechanical description of this sequence (Feynman and Vernon 1963), although probability amplitudes are required, and the effect on the probability distribution is not as dramatic.

The ingredients in the expression for $P(\mathbf{x}(2)|\mathbf{y}(1))$, Eq. (3.29), are identities on conditional probabilities and the assumption of the Markov property for the propagation of the state between measurements. No restriction was required on the quality of the measurements or the distribution of errors in the measurements, nor did we have to specify the model errors or noise sources in the dynamics that enter $P(\mathbf{x}(n+1)|\mathbf{x}(n))$. We will, naturally, be required to do all of that when we seek to numerically evaluate such expressions.

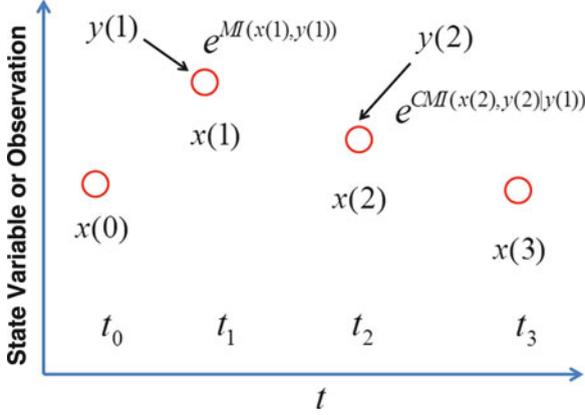


Fig. 3.3 Illustration of a measurement at time t_1 and again at t_2 of the L quantities $\mathbf{y}(1)$ and $\mathbf{y}(2)$ as the state of a model moves from $t_0 \rightarrow t_1 \rightarrow t_2 \rightarrow t_3$. The result of the measurement at t_1 is the addition of a factor $\exp[\text{MI}(\mathbf{x}(1), \mathbf{y}(1))]$ to the Markov chain result when no measurements are made. The result of the measurement at t_2 is the addition of a factor $\exp[\text{CMI}(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1))]$ to the Markov chain result when no measurements are made. The factor $\text{MI}(\mathbf{x}(1), \mathbf{y}(1))$ is here the same as $\text{CMI}(\mathbf{x}(1), \mathbf{y}(1)|\mathbf{y}(0))$ as no measurement is made at t_0

3.2.3 Two Measurements

If we make two measurements at times t_1 and t_2 and wish to know the distribution at time $t_3 \geq t_2 \geq t_1$ conditioned upon these measurements, we seek an expression for

$$P(\mathbf{x}(3)|\mathbf{y}(2), \mathbf{y}(1)). \quad (3.33)$$

We expect that we must have propagators (transition probabilities) $P(\mathbf{x}(1)|\mathbf{x}(0))$ from t_0 to t_1 , $P(\mathbf{x}(2)|\mathbf{x}(1))$ from t_1 to t_2 , and $P(\mathbf{x}(3)|\mathbf{x}(2))$ from t_2 to t_3 and some additional factors associated with the measurements along the path (Fig. 3.3). For this we first use the Markov property and the resulting Chapman–Kolmogorov equation of the dynamics to write

$$P(\mathbf{x}(3)|\mathbf{y}(2), \mathbf{y}(1)) = \int d^D x(2) P(\mathbf{x}(3)|\mathbf{x}(2)) P(\mathbf{x}(2)|\mathbf{y}(2), \mathbf{y}(1)), \quad (3.34)$$

then from the definition of conditional probabilities, we have

$$\begin{aligned} P(\mathbf{x}(2)|\mathbf{y}(2), \mathbf{y}(1)) &= \frac{P(\mathbf{x}(2), \mathbf{y}(2), \mathbf{y}(1))}{P(\mathbf{y}(2), \mathbf{y}(1))} \\ &= \frac{P(\mathbf{x}(2), \mathbf{y}(2), \mathbf{y}(1))}{P(\mathbf{y}(2), \mathbf{y}(1)) P(\mathbf{x}(2)|\mathbf{y}(1))} P(\mathbf{x}(2)|\mathbf{y}(1)) \\ &= \frac{P(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1))}{P(\mathbf{y}(2)|\mathbf{y}(1)) P(\mathbf{x}(2)|\mathbf{y}(1))} P(\mathbf{x}(2)|\mathbf{y}(1)). \end{aligned} \quad (3.35)$$

From the result with one measurement, we then may write

$$\begin{aligned}
& P(\mathbf{x}(3)|y(2), \mathbf{y}(1)) \\
&= \int d^D x(2) d^D x(1) d^D x(0) P(\mathbf{x}(3)|\mathbf{x}(2)) \\
&\quad \left\{ \frac{P(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1))}{P(\mathbf{y}(2)|\mathbf{y}(1)) P(\mathbf{x}(2)|\mathbf{y}(1))} \right\} P(\mathbf{x}(2)|\mathbf{x}(1)) \exp[\text{MI}(\mathbf{x}(1), \mathbf{y}(1))] \\
&\quad P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)), \tag{3.36}
\end{aligned}$$

or

$$\begin{aligned}
& P(\mathbf{x}(3)|y(2), \mathbf{y}(1)) \\
&= \int d^D x(2) d^D x(1) d^D x(0) P(\mathbf{x}(3)|\mathbf{x}(2)) \\
&\quad \exp[\text{CMI}(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1))] P(\mathbf{x}(2)|\mathbf{x}(1)) \exp[\text{MI}(\mathbf{x}(1), \mathbf{y}(1))] \\
&\quad P(\mathbf{x}(1)|\mathbf{x}(0)) P(\mathbf{x}(0)), \tag{3.37}
\end{aligned}$$

and the conditional mutual information, of the D -dimensional state $\mathbf{x}(2)$, and the L -dimensional measurement, conditioned on the L -dimensional measurement $\mathbf{y}(1)$, is given as

$$\text{CMI}(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1)) = \log \left\{ \frac{P(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1))}{P(\mathbf{y}(2)|\mathbf{y}(1)) P(\mathbf{x}(2)|\mathbf{y}(1))} \right\}. \tag{3.38}$$

In this expression we could have written

$$\text{MI}(\mathbf{x}(1), \mathbf{y}(1)) = \text{CMI}(\mathbf{x}(1), \mathbf{y}(1)|\mathbf{y}(0)), \tag{3.39}$$

as there is no measurement at t_0 on which to condition the mutual information.

As with a single measurement, two measurements transfer information from the observations $\mathbf{y}(1)$, $\mathbf{y}(2)$ to the model dynamics constraining the allowed model state space trajectory. The action now has the form, with $\mathbf{X} = \{\mathbf{x}(3), \mathbf{x}(2), \mathbf{x}(1), \mathbf{x}(0)\}$,

$$\begin{aligned}
A_0(\mathbf{X}, \mathbf{y}(1), \mathbf{y}(2)) &= -\text{CMI}(\mathbf{x}(1), \mathbf{y}(1)|\mathbf{y}(0)) - \text{CMI}(\mathbf{x}(2), \mathbf{y}(2)|\mathbf{y}(1), \mathbf{y}(0)) \\
&\quad - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))]. \tag{3.40}
\end{aligned}$$

Any function on the path \mathbf{X} has a conditional expectation value

$$E[G(\mathbf{X})|y(1), \mathbf{y}(2)] = \frac{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{y}(1), \mathbf{y}(2))] G(\mathbf{X})}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{y}(1), \mathbf{y}(2))]} \tag{3.41}$$

As with one measurement, when one uses the action in evaluating the expected value of a quantity along the path \mathbf{X} , there is a useful simplification arising from factors in $A_0(\mathbf{X}, \mathbf{y}(1), \mathbf{y}(2))$ that are independent of \mathbf{X} . Dropping these \mathbf{X} independent factors as they cancel between the numerator and denominator, the effective form of the action is then

$$A_0(\mathbf{X}, \mathbf{y}(1), \mathbf{y}(2))_{\text{effective in Eq. (3.41)}} = -\log[P(\mathbf{y}(2)|\mathbf{x}(2), \mathbf{y}(1))] - \log[P(\mathbf{y}(1)|\mathbf{x}(1))] \\ - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))]. \quad (3.42)$$

“Collapse” of the Probability Distribution

It may be interesting to draw an analogy here to the observation process in quantum theory (Auletta et al. 2009). Of course, in quantum mechanics one is discussing complex probability amplitudes rather than real probability distributions. When an observation is made in quantum theory, the eigenvalue of an Hermitian operator is measured, and this informs the quantum system which eigenvalue of that operator the wave function carries at the time of observation. In a sense that can be made precise, information flows from the measurement to the future description of the wave function. If the eigenvalue is discrete, then there is really no classical analogy. However, if the quantum mechanical measurement is from a continuum, such as position or momentum, then in the absence of noise, the wave function “collapses”, and the analog here is that the classical probability distribution in state space is constrained within the observed subspace. This is a direct analogy with the collapse of the distribution, especially if the dynamics is chaotic.

When orbits of the dynamical system are chaotic, small errors in the location of any state at some time t will lead to large deviations at later times. This can lead, absent observations, to the state of the system moving as far away as the “size” of the attractor. If one has an observation at a time $t_{\text{obs}} > t$, then through the conditional mutual information term in the path integral, information about the appropriate location of the orbit is transferred to the dynamics, and the orbit is moved, in the sense of the mass of a probability distribution toward the location observed. In our general treatment we have allowed that the observation may well be of an L -dimensional subset of the state variables, suggesting a sequence of observations, rather than a single measurement, may be required to significantly influence the orbit. Whether this occurs is a quantitative question of the magnitude of the information provided by the measurement compared to the entropy generated by the orbit as the dynamics moves forward on its own.

3.3 The General Data Assimilation Problem

3.3.1 Differential Equations to Discrete Time Maps

Now we turn to the general problem of utilizing information from observed data for the purpose of completing a model we have made of those observations. The model, constructed on physical or biophysical principles, is comprised of two parts:

- A set of differential equations, stochastic or deterministic, ordinary or partial, which gives the time dependence or dynamical rule on moving the D -dimensional model state $\mathbf{x}(t)$ forward from some time to a future time. Discrete time dynamics, usually arising from a numerical form for solving the differential equations, will appear at this step.
- A measurement rule that specifies how the L -dimensional observed quantities $\mathbf{y}(t)$ are related to the D -dimensional state variables $\mathbf{x}(t)$ of the model. If the measurements made are related to the state of the system $\mathbf{x}(t)$ as $h_l(\mathbf{x}(t))$; $l = 1, 2, \dots, L$, then we would seek to provide information from those measurements to the model such that $y_l(t) \approx h_l(\mathbf{x}(t))$.

In each of these parts, one encounters unknown parameters, and the information in the data must also be used to estimate these.

We begin with a differential equation or discrete time map for the model dynamics. This is not formulated by any of the methods we discuss in this book, but is constructed using physical and biophysical principles known to us about the source of our data. The methods we attend to here require data related to the model dynamics through measurement functions and a model constructed as a sound guess based on physical ideas about the source of the data. The data assimilation problem is to complete the model by specifying estimates for all unknown fixed parameters in the model and the measurement functions as well as estimates of all the unobserved model state variables at some time T . We will usually take that time to be at the end of an observation window in time.

As addressed earlier, in practice we discretize space \mathbf{r} and time if we begin with PDEs. Discretizing space and collecting the degrees of freedom at each grid point (or equivalent spatial label) along with the vector index α into a label $a = 1, 2, \dots, D$, we arrive at a set of D ordinary differential equations for $\mathbf{x}(t) = \{x_1(t), x_2(t), \dots, x_D(t)\}$

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t), \mathbf{peq}), \quad (3.43)$$

where \mathbf{peq} is also a set of time-independent parameters.

Now we specify a discrete set of times $\{t_0, t_1, \dots, t_m = T\}$ at each of which measurements of L quantities $\mathbf{y}(t_n) = \{y_1(t_n), y_2(t_n), \dots, y_L(t_n)\}$ could be made. In the absence of errors in the measurements, we associate the measurements with L functions of the state variables $y_l(t_n) = h_l(\mathbf{x}(t_n), \mathbf{pmeas})$ where the \mathbf{pmeas} are

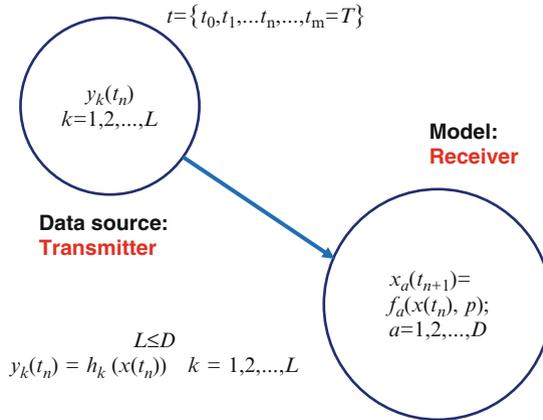


Fig. 3.4 Illustration of the data assimilation problem. At each time in an observation period $\{t_0, t_1, \dots, t_m = T\}$, L measurements $y_l(t_n)$; $l = 1, 2, \dots, L$ are made. A dynamical model in discrete time taking $\mathbf{x}(t_n) \rightarrow \mathbf{x}(t_{n+1})$ for the D -dimensional state variables $\mathbf{x}(t_n) = [x_1(t_n), x_2(t_n), \dots, x_D(t_n)]$; $n = 0, 1, 2, \dots, m$ is created, satisfying $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$; $a = 1, 2, \dots, D$. If the connection between $\mathbf{x}(t_n)$ and $\mathbf{x}(t_{n+1})$ is explicit, we have $\mathbf{x}(t_{n+1}) = \mathbf{f}(\mathbf{x}(t_n), \mathbf{p})$. The \mathbf{p} are time-independent parameters in the model. A measurement model indicating the nonlinear function of the state variables $h_l(\mathbf{x}(t), \mathbf{p}_{\text{meas}})$ corresponding to the observed $y_l(t)$ is developed. Information is passed from the observations to the model with the goal of estimating all the fixed parameters \mathbf{p} and \mathbf{p}_{meas} as well as the $D - L$ unobserved state variables among the $\mathbf{x}(t)$. When the observation period is completed, knowledge of the whole set of $\mathbf{x}(T)$ and the fixed parameters may be used to predict into $t > T$ fixed parameters \mathbf{p} and \mathbf{p}_{meas} as well as the $D - L$ unobserved state variables among the $\mathbf{x}(t)$. When the observation period is completed, knowledge of the whole set of $\mathbf{x}(T)$ and the fixed parameters may be used to predict into $t > T$ using $\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$

fixed constants in the specification of the measurement functions. Discretizing the continuous time ordinary differential equations, Eq.(3.43), by giving a rule that takes the state at t_n $\mathbf{x}(t_n)$ to the state at t_{n+1} : $g_a(\mathbf{x}(n), \mathbf{x}(t_{n+1}), \mathbf{p}) = 0$; $\mathbf{p} = \{p_1, p_2, \dots, p_{NP}\}$. All fixed parameters, including those in the rule for discretizing the time-stepping rule, are in \mathbf{p} . If the rule for proceeding forward in discrete time is explicit, we write it as $\mathbf{x}(t_{n+1}) = \mathbf{f}(\mathbf{x}(t_n), \mathbf{p})$ or $\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ (Fig. 3.4).

Now our problem is specified:

1. Measure L quantities at any (or all) of the $m + 1$ time locations t_n ; $n = 0, 1, \dots, m$ in an observation window: $\{t_0, t_1, \dots, t_m\}$.
2. Specify some model dynamics taking the state $\mathbf{x}(t_n)$ to the state $\mathbf{x}(t_{n+1})$ through that time interval. This is done by giving the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ which, in a deterministic model, is $\delta^D(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}))$. This is the step where the physical properties of the measurements are considered.
3. Specify the connection between the observed quantities $y_l(t)$ and the measurement function of the model state variables $h_l(\mathbf{x}(n), \mathbf{p}_{\text{meas}})$.

We wish to use the information in the $L(m + 1)$ observations to estimate the $D(m + 1)$ observed and unobserved state variables through the observation window. Use this information also to estimate the NP unknown time-independent parameters in the dynamical model and the measurement functions. As $L \leq D$ and usually is much less than D , we have no guarantee that there is enough information in the measurements to allow accurate estimation of the states, L observed $D - L$ unobserved, and of the parameters of the model plus the measurement functions.

When this is accomplished we will have two types of important information about the model of the underlying physical dynamics:

- We will have a quantitative picture of the state of the experimental or field system at all times in the observation window. This can be assessed for its physically realistic and relevant structure. To make this assessment requires knowledge of the physical system, and this is not discussed here. Usually this is specific to the physics of the observed system.
- We will have an estimate of the unknown time-independent parameters of the observed dynamics to use in assessing properties of the system.
- We will have an estimate of the full state and the full set of fixed parameters of our model of the observed system. This allows us to use the completed model to predict the future behavior of the observed dynamics.

3.3.2 Errors and Noise: Stochastic Data Assimilation

The idealized situation just outlined might seem to imply we have a perfect knowledge of the dynamical model, noise-free measurements $y_l(t_n)$ to compare with functions $h_l(\mathbf{x}(t_n, \mathbf{p}_{\text{meas}}))$ of the model state variables and accurate knowledge of the state of the model when observations begin: $\mathbf{x}(t_0)$. In fact, we typically have none of these. The measurements are inaccurate because of noise in the environment, the instruments, and elsewhere; the model is in error as we do not, in fact, know the detailed structure of the dynamics controlling the generation of the data, and we typically have inaccurate knowledge of the state of the system $\mathbf{x}(t_0)$ when observations begin (Fig. 3.5).

Our attention is thus directed to a probabilistic description of the data assimilation process within which distribution functions of the states of the model are the focus. The essential quantity of interest is the $(m + 1)(D + L) + NP$ -dimensional distribution function of the $(m + 1)D$ states in the time interval $[t_0, t_m = T]$, the $(m + 1)L$ observations in the same time window, and the NP unknown time-independent parameters. We'll call this $P(\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_m), \mathbf{y}(t_0), \mathbf{y}(t_1), \dots, \mathbf{y}(t_m))$ where we do not explicitly show the parameter dependence of this joint probability distribution. This is a distribution over an often enormously high-dimensional space and is the core of our inquiries. We look now at the slightly more restricted quantity identified as the probability distribution of the state $\mathbf{x}(t_m)$ at the end of the observation interval conditioned on the measurements up

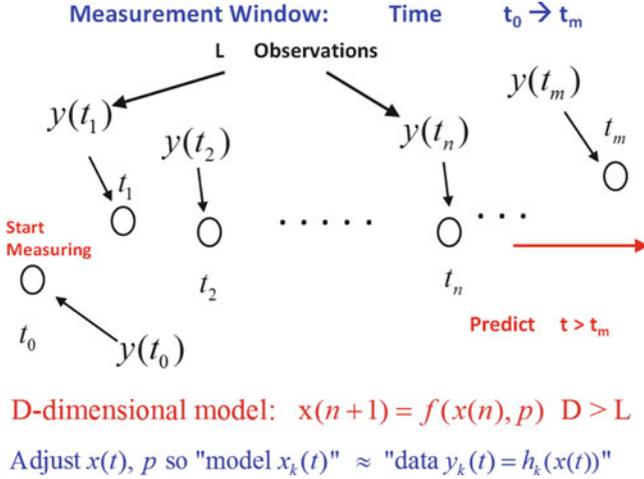


Fig. 3.5 A schematic of the acquisition of information from L -dimensional measurements at times t_n ; $n = 0, 1, \dots, m$. At each measurement time, the conditional probability for the state $\mathbf{x}(t_n)$ is modified by a factor involving the conditional mutual information of the state $\mathbf{x}(t_n)$ and the measurement $\mathbf{y}(t_n)$ conditioned on the previous measurements. If a measurement is not made, no such factor is present. In this graphic we compare the measurement with the function of the model output to which the measurement corresponds

to that time $t_m = T$: $P(\mathbf{x}(t_m)|\mathbf{Y}(m)) = P(\mathbf{x}(m)|\mathbf{Y}(m))$, where $\mathbf{Y}(m) = \{\mathbf{y}(t_m), \mathbf{y}(t_{m-1}), \dots, \mathbf{y}(t_1), \mathbf{y}(t_0)\} = \{\mathbf{y}(m), \mathbf{y}(m-1), \dots, \mathbf{y}(0)\}$ and

$$P(\mathbf{x}(m)|\mathbf{Y}(m)) = \frac{P(\mathbf{x}(m), \mathbf{Y}(m))}{P(\mathbf{Y}(m))}. \tag{3.44}$$

The conditional probability distribution $P(A|B)$ conveys a quantitative measure of the information one learns about measurements of a set of quantities A when we have knowledge of another set of quantities B .

- First, if the sets A and B are independent, $P(A, B) = P(A)P(B)$, then $P(A|B) = P(A)$, and nothing has changed our knowledge of the set A by the measurements of elements of the set B .
- Second, recalling that the overall information in a set A is Shannon's entropy (Fano 1961)

$$H(A) = - \sum_A P(A) \log\{P(A)\}, \tag{3.45}$$

we see that

$$\begin{aligned}
 H(A|B) &= - \sum_{A,B} P(A, B) \log\{P(A, B)\} - \sum_{A,B} P(A, B) \log\{P(B)\} \\
 &= - \sum_{A,B} P(A, B) \log\{P(A, B)\} - \sum_B P(B) \log\{P(B)\} \\
 &= H(A, B) - H(B),
 \end{aligned} \tag{3.46}$$

or the information in the conditional probability is the information in the joint distribution of A and B reduced by the knowledge already acquired about B . For $P(\mathbf{x}(m)|\mathbf{Y}(m))$ this tells us that the information we have about the state of the model system at time t_m is the information about the combined state and previous observations $\mathbf{Y}(m)$ minus the information provided by the observations. The increase in knowledge is due to the state moving from time t_{m-1} to t_m and exploring state space over a time $t_m - t_{m-1}$ and the knowledge gained by the observation at t_m .

Pesin's Theorem

The connection of the information about a set A and the dynamics of nonlinear systems are found in Pesin's theorem (Pesin 1977) which relates the entropy in the movement of the state of a dynamical system on an attractor to the positive Lyapunov exponents $\lambda_a > 0$ for the time development of perturbations to a system orbit. This relationship says the entropy, in bits per second, is given as

$$H(\text{orbits on an attractor}) = \sum_a \lambda_a; \quad \lambda_a > 0, \tag{3.47}$$

and this has the intuitive sense that positive Lyapunov exponents, associated as they are with instability and the departure of one orbit of a nonlinear system from a perturbed orbit, reveals information about where the system lies in state space (Abarbanel 1996).

The qualitative idea is that in physical systems, we always have a finite resolution ball within which at some time t_0 we cannot distinguish two or more distinct states, say, $\mathbf{x}^{(1)}(t_0)$ and $\mathbf{x}^{(2)}(t_0)$. With some sense of distance giving the resolution ball a radius R_{res} : $|\mathbf{x}^{(1)}(t_0) - \mathbf{x}^{(2)}(t_0)| < R_{\text{res}}$. If the dynamics is chaotic, this distance grows as $\exp \lambda(t - t_0)$ with $\lambda > 0$. At some time greater than t_0 , the distance will be larger than $2R_{\text{res}}$, and the two resolution balls about each state space point will be nonoverlapping thus allowing one to distinguish between the two points. This reveals information about the dynamical system not observable at t_0 because of the finite resolution.

The message of Pesin's theorem is that there is a quantitative measure of the information generated by a chaotic trajectory in a nonlinear system. The information transferred from data to a dynamical model which has positive Lyapunov exponents

must be equal or greater than the entropy generated by the system as it expresses the realization of local instabilities giving rise to these positive Lyapunov exponents. In the general formulation presented in the chapter we have a precise way to evaluate the information in bits transferred along the model trajectory within the observation window:

$$-\sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))]. \quad (3.48)$$

If we can transfer, through this part of the action $A_0(\mathbf{X}, \mathbf{Y})$, more than the uncertainty (entropy) generated by the dynamical instabilities of the model, then we stand in a good place to provide accurate estimates of the full system state $\mathbf{x}(t_m = T)$ at the end of observations. We do not have a precise formula for how many observations that requires at each observation time nor does it tell us which observations are to be made. We explore this in a numerical, somewhat cut-and-try manner in a later chapter.

General Result

We want to express the conditional probability distribution $P(\mathbf{x}(t_m)|\mathbf{Y}(m)) = P(\mathbf{x}(m)|\mathbf{Y}(m))$ in terms of the same quantity one step earlier in time $P(\mathbf{x}(t_{m-1})|\mathbf{Y}(m-1)) = P(\mathbf{x}(m-1)|\mathbf{Y}(m-1))$. Using the definition of the conditional probability we write

$$\begin{aligned} P(\mathbf{x}(m)|\mathbf{Y}(m)) &= \frac{P(\mathbf{x}(m), \mathbf{y}(m), \mathbf{Y}(m-1))}{P(\mathbf{y}(m), \mathbf{Y}(m-1))} \\ &= \frac{P(\mathbf{x}(m), \mathbf{y}(m), \mathbf{Y}(m-1))/P(\mathbf{Y}(m-1))}{P(\mathbf{y}(m), \mathbf{Y}(m-1))/P(\mathbf{Y}(m-1))} \\ &= \frac{P(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1))}{P(\mathbf{y}(m)|\mathbf{Y}(m-1))} \\ &= \frac{P(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1))}{P(\mathbf{y}(m)|\mathbf{Y}(m-1))} \frac{P(\mathbf{x}(m)|\mathbf{Y}(m-1))}{P(\mathbf{x}(m)|\mathbf{Y}(m-1))} \\ &= \frac{P(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1))}{P(\mathbf{y}(m)|\mathbf{Y}(m-1)) P(\mathbf{x}(m)|\mathbf{Y}(m-1))} P(\mathbf{x}(m)|\mathbf{Y}(m-1)) \\ &= \exp[\text{CMI}(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1))] P(\mathbf{x}(m)|\mathbf{Y}(m-1)). \quad (3.49) \end{aligned}$$

In writing this we have abbreviated $\mathbf{x}(t_n) = \mathbf{x}(n)$ and $\mathbf{y}(t_n) = \mathbf{y}(n)$ and recognized the conditional mutual information

$$\text{CMI}(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1)) = \log \left\{ \frac{P(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1))}{P(\mathbf{x}(n)|\mathbf{Y}(n-1)) P(\mathbf{y}(n)|\mathbf{Y}(n-1))} \right\} \quad (3.50)$$

between the L -dimensional observation $\mathbf{y}(n)$ and the D -dimensional model state $\mathbf{x}(n)$ conditioned on the measurements up to t_{n-1} ; $\mathbf{Y}(n-1) = \{\mathbf{y}(n-1), \mathbf{y}(n-2), \dots, \mathbf{y}(0)\}$. This answers the question: how much information (in bits if the logarithm is taken in base 2) do we learn about the state $\mathbf{x}(n)$ by measuring $\mathbf{y}(n)$ given the n earlier measurements $\{\mathbf{y}(n-1), \mathbf{y}(n-2), \dots, \mathbf{y}(0)\}$?

We have stated the model associated with our observed data as a rule taking the state at time t_n to the state one time step ahead

$$g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0; \quad a = 1, 2, \dots, D, \quad (3.51)$$

making the assumption that the state $\mathbf{x}(n+1)$ depends only on the previous state $\mathbf{x}(n)$ and some fixed parameters. This is the key property of a Markov model, and a general property of these models is the Chapman–Kolmogorov relation. This comes from the following identity:

$$\begin{aligned} P(x(m)|\mathbf{Y}(m-1)) &= \int d^D x(m-1) P(\mathbf{x}(m), \mathbf{x}(m-1)|\mathbf{Y}(m-1)) \\ &= \int d^D x(m-1) \frac{P(\mathbf{x}(m), \mathbf{x}(m-1), \mathbf{Y}(m-1))}{P(\mathbf{Y}(m-1))} \\ &= \int d^D x(m-1) \frac{P(\mathbf{x}(m), \mathbf{x}(m-1), \mathbf{Y}(m-1))}{P(\mathbf{x}(m-1), \mathbf{Y}(m-1))} \frac{P(\mathbf{x}(m-1), \mathbf{Y}(m-1))}{P(\mathbf{Y}(m-1))} \\ &= \int d^D x(m-1) P(\mathbf{x}(m)|\mathbf{x}(m-1), \mathbf{Y}(m-1)) P(\mathbf{x}(m-1)|\mathbf{Y}(m-1)), \end{aligned}$$

and since the dynamics is Markov, then

$$P(\mathbf{x}(m)|\mathbf{x}(m-1), \mathbf{Y}(m-1)) = P(\mathbf{x}(m)|\mathbf{x}(m-1)) \quad (3.52)$$

and

$$P(\mathbf{x}(m)|\mathbf{Y}(m-1)) = \int d^D x(m-1) P(\mathbf{x}(m)|\mathbf{x}(m-1)) P(\mathbf{x}(m-1)|\mathbf{Y}(m-1)). \quad (3.53)$$

The transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ expresses the underlying dynamics of the model processes taking the state $\mathbf{x}(n)$ to the state $\mathbf{x}(n+1)$. It is a $D(m+1) \times D(m+1)$ matrix in the space of states $\mathbf{x}(n)$. If the process $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1)$ is deterministic,

$$P(\mathbf{x}(n+1)|\mathbf{x}(n)) = \delta^D(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p})). \quad (3.54)$$

Combining Eq. (3.53) with the identities on conditional probabilities, we arrive at the key recursion relation

$$\begin{aligned}
P(\mathbf{x}(m)|\mathbf{Y}(m)) &= \exp[\text{CMI}(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1))] \\
&\int d^D x(m-1) P(\mathbf{x}(m)|\mathbf{x}(m-1)) P(\mathbf{x}(m-1)|\mathbf{Y}(m-1)),
\end{aligned} \tag{3.55}$$

which, applied along a Markov chain of states from the final time t_m back to the time observations are begun, results in

$$\begin{aligned}
P(\mathbf{x}(m)|\mathbf{Y}(m)) &= \int \prod_{n=0}^{m-1} d^D x(n) \exp \left[\sum_{n=0}^m \text{CMI}(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1)) \right] \\
&\prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)) P(\mathbf{x}(0)) \\
&= \int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))],
\end{aligned} \tag{3.56}$$

where $\mathbf{X} = \{\mathbf{x}(m), \mathbf{x}(m-1), \dots, \mathbf{x}(0)\}$, $\mathbf{Y}(m) = \{\mathbf{y}(m), \mathbf{y}(m-1), \dots, \mathbf{y}(0)\}$ and

$$\begin{aligned}
A_0(\mathbf{X}, \mathbf{Y}(m)) &= - \sum_{n=0}^m \text{CMI}(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1)) \\
&- \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))].
\end{aligned} \tag{3.57}$$

The conditional expected value of any quantity $G(\mathbf{X})$ along the path \mathbf{X} is given as

$$E[G(\mathbf{X})|\mathbf{Y}] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}, \tag{3.58}$$

and evaluating this integral with various choices for $G(\mathbf{X})$ will allow us to address all questions of interest in statistical data assimilation.

The expected path through the space of paths during the observation window $\{t_0, t_1, \dots, t_m = T\}$ comes from selecting $G(\mathbf{X}) = \mathbf{X}$. The computation of the RMS variation about this expected path is performed with the information contained in $G(\mathbf{X}) = \{x_a(0)^2, x_a(1)^2, \dots, x_a(m)^2\}$. Other moments about the expected path are evaluated in the same way using different powers of the $x_a(n)$. The marginal distribution at time t_j $P_{x_c(j)}(z)$ for some component $x_c(j)$ comes from selecting $G(\mathbf{X}) = \delta(z - x_c(j))$.

When used in evaluating Eq. (3.58) there is, as pointed out earlier, a simplification in the form of the measurement term in the action. To identify this we note

$$\begin{aligned}
& \frac{P(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1))}{P(\mathbf{x}(n)|\mathbf{Y}(n-1)) P(\mathbf{y}(n)|\mathbf{Y}(n-1))} \\
&= \frac{P(\mathbf{x}(n), \mathbf{y}(n), \mathbf{Y}(n-1))}{P(\mathbf{x}(n), \mathbf{Y}(n-1))} P(\mathbf{y}(n)|\mathbf{Y}(n-1))^{-1} \\
&= P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1)) P(\mathbf{y}(n)|\mathbf{Y}(n-1))^{-1}, \tag{3.59}
\end{aligned}$$

and the final factor does not depend on the path \mathbf{X} over which we integrate. This then cancels between the numerator and the denominator, leaving us the residual action when evaluating the expectation values of function on the path

$$\begin{aligned}
A_0(\mathbf{X}, \mathbf{Y}(m))_{\text{residual for Eq. (3.58)}} &= - \sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))] \\
&\quad - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))], \tag{3.60}
\end{aligned}$$

and the expected value of any function $G(\mathbf{X})$ on the path is

$$E[G(\mathbf{X})|\mathbf{Y}] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}]}. \tag{3.61}$$

3.4 Approximating the Action

The general expression for the expected value along a path Eq.(3.61) becomes useful only when we specify the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ and the model $\mathbf{g}(\mathbf{x}(n), \mathbf{x}(n+1), \mathbf{p})$ as well as the measurement function relating observations $\mathbf{y}(n)$ to model output $\mathbf{x}(n)$.

We start with the transition probability which we recall is

$$\delta^D(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})) \tag{3.62}$$

when the propagation of the state in time is deterministic. When there is model error, associated with fluctuations in the environment of the dynamical states $\mathbf{x}(n)$ not explicitly accounted for in the dynamical equations or associated with the absence of all of the precise terms in the dynamical equations, we expect this error to appear at minimum as reduced resolution in our ability to establish the present location of the state. If the distribution $\delta(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}))$ is replaced by the smoother, lower resolution $P_{\ominus}(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}))$, then some parameters in this distribution function reflect that loss of resolution. If the model errors are given by a Gaussian, then we could write

$$P_{\Theta}(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})) = \exp \left[- \sum_{a,b=1}^D g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \frac{R_{ab}(n)}{2} g_b(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \right], \quad (3.63)$$

in an unnormalized format. The normalization drops out of the integral Eq. (3.61). By no means are we restricted to using Gaussian distributions here.

This effective form of the action Eq.(3.60) makes clear that in specifying the relation between observations $\mathbf{y}(n)$ and model output $\mathbf{x}(n)$, we are asked to represent $P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))$. If the measurements at some t_n are independent of measurements made earlier than t_n , the conditional mutual information term becomes

$$P(\mathbf{y}(n)|\mathbf{x}(n)), \quad (3.64)$$

and if the measurement function relates model output $\mathbf{x}(n)$ to observations $\mathbf{y}(n)$ by an additive noise term

$$y_l(n) = x_l(n) + \theta_l(n), \quad l = 1, 2, \dots, L, \quad (3.65)$$

then

$$P(\mathbf{y}(n)|\mathbf{x}(n)) = P_{\Theta}(\mathbf{y}(n) - \mathbf{x}(n)). \quad (3.66)$$

This relationship holds for the components of the model state $\mathbf{x}(n)$ that are observed. If the observations $y_l(n)$ correspond to richer measurement functions $h_l(\mathbf{x}(n))$, then we will need $P(\mathbf{y}(n)|\mathbf{x}(n)) = P_{\Theta}(\mathbf{y}(n) - \mathbf{h}(\mathbf{x}(n)))$ here.

Further, if the noise in the independent measurements is Gaussian distributed, we have (again unnormalized)

$$\text{CMI}(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1)) = - \sum_{l,k=1}^L (h_l(\mathbf{x}(n)) - y_l(n)) \frac{[R_m(n)]_{lk}}{2} (h_k(\mathbf{x}(n)) - y_k(n)), \quad (3.67)$$

and should there be no measurement at t_n , we take the $L \times L$ matrix $R_m(n)$ to be zero.

The structure of the action indicates how to proceed when the observations at two time t_n and t_{n-1} , say, are not independent of each other. This can be important in application (Hamill 2006), though it is hard to find a discussion of this matter elsewhere in the literature.

For example, suppose the measurement model relates a linear combination of the measurements $\mathbf{y}(n)$ and $\mathbf{y}(n-1)$ to the function of the state variable $\mathbf{x}(n)$, in the following manner

$$x_l(n) = Ay_l(n) + (1-A)y_l(n-1) + \theta_l(n). \quad (3.68)$$

Also we assume $y(n)$ does not depend on $\mathbf{y}(n-2), \mathbf{y}(n-3), \dots$. Then we may write

$$y_l(n) = \frac{x_l(n)}{A} - \left(\frac{1}{A} - 1\right) y_l(n-1) + \frac{1}{A} \theta_l(n); \quad (3.69)$$

$0 \leq A \leq 1$. This means $P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{y}(n-1))$ is

$$P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{y}(n-1)) = P_{\Theta/A} \left(\mathbf{y}(n) - \frac{\mathbf{x}(n)}{A} + \left(\frac{1}{A} - 1\right) \mathbf{y}(n-1) \right). \quad (3.70)$$

3.5 The Value of a Measurement

The path integral structure allows us to answer the interesting question: what is the value of one additional measurement to the receiving model? If we look at Eq. (3.49), divide by $P(\mathbf{x}(m))$, and take the logarithm of the result, we arrive at

$$\begin{aligned} \log \left\{ \frac{P(\mathbf{x}(m), \mathbf{Y}(m))}{P(\mathbf{x}(m)) P(\mathbf{Y}(m))} \right\} &= \text{CMI}(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1)) \\ &+ \log \left\{ \frac{P(\mathbf{x}(m), \mathbf{Y}(m-1))}{P(\mathbf{x}(m)) P(\mathbf{Y}(m-1))} \right\}; \end{aligned} \quad (3.71)$$

at each measurement time t_n , the additional information is $\text{CMI}(\mathbf{x}(n), \mathbf{y}(n)|\mathbf{Y}(n-1))$.

Now integrate this with $P(\mathbf{x}(m), \mathbf{Y}(m))$. On the left-hand side we have the average mutual information when $m+1$ measurements $\mathbf{Y}(m)$ are transmitted to the model and the state $\mathbf{x}(m)$ occurs. On the right-hand side, we have the term

$$\begin{aligned} &\int d^D x(m) \prod_{n=0}^m d^L y(n) P(\mathbf{x}(m), \mathbf{Y}(m)) \log \left\{ \frac{P(\mathbf{x}(m), \mathbf{Y}(m-1))}{P(\mathbf{x}(m)) P(\mathbf{Y}(m-1))} \right\} \\ &= \int d^D x(m) \prod_{n=0}^{m-1} d^L y(n) P(\mathbf{x}(m), \mathbf{Y}(m-1)) \log \left\{ \frac{P(\mathbf{x}(m), \mathbf{Y}(m-1))}{P(\mathbf{x}(m)) P(\mathbf{Y}(m-1))} \right\}. \end{aligned} \quad (3.72)$$

This term is the average mutual information when m measurements $\mathbf{Y}(m-1)$ are transmitted to the model and the state $\mathbf{x}(m)$ occurs.

The difference between these

$$\begin{aligned} &\text{AMI}(\mathbf{x}(m) \text{ occurs}; m+1 \text{ measurements}) - \text{AMI}(\mathbf{x}(m) \text{ occurs}; m \text{ measurements}) \\ &= \int d^D x(m) \prod_{n=0}^m d^L y(n) P(\mathbf{x}(m), \mathbf{Y}(m)) \text{CMI}(\mathbf{x}(m), \mathbf{y}(m)|\mathbf{Y}(m-1)), \end{aligned} \quad (3.73)$$

which is the average conditional mutual information conveyed by the $(m + 1)$ st measurement $\mathbf{y}(m)$ from the transmitter (the data generating process) to the receiver (the model). This tells us on the average how much information is gained by the model via one additional measurement. If the logarithm is taken to base 2, this quantity is in bits.

This permits the assessment of the value in bits of an additional measurement. It depends on the measurements and on the model, and it tells us in the context of a particular model for the processes generating the data whether the additional measurement has enough value to do the work associated with making the observation. If measurements are difficult or expensive yet yield a large information flow, they may be worth pursuing. If the opposite is true, one knows from this calculation what is being missed.

3.6 Predicting

In a broader situation than identified until now, one may have a combination of observation windows of various temporal lengths, interspersed with windows where no observations are made. In the latter, the conditional mutual information terms are absent, and we utilize the results from the first part of this chapter to move the dynamical system forward between observation windows. Both in the observation window and outside of it, one may use different time steps for moving the model forward in time than those associated with the observation times t_n . This may be of some value when the discrete time step algorithm $\mathbf{x}(n + 1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ requires smaller time steps to converge well.

Using an observation window we can “complete” the model instantiated in the transition probability $P(\mathbf{x}(n + 1)|\mathbf{x}(n))$ by determining the expected model states $\mathbf{x}(t)$ and all the model and measurement parameters \mathbf{p} from information provided by the observations. Putting the fixed parameters on the same footing as the states by providing $d\mathbf{p}(t)/dt = 0$ as a model equation is a device that works. The expected values of functions on the path and moments about them as well as the full marginal distribution for these quantities come from evaluating the path integral. Predictions for the behavior of the completed model for times outside the observation windows are now a matter of evaluations of the same category of path integral already encountered in an observation window but, as in the beginning of this chapter, without further observations.

We will return to predicting as we examine the various methods for evaluating the path integrals in the coming chapters.

There is a circumstance we know in which predicting may not be of interest to the user of statistical data assimilation. Suppose one has observations of a dynamical system for which one has a well-established, well-known model. In the model there may be a few unknown parameters, but the important unknown quantities are values of the state variables hard or not possible to observe. Then under the assumption that enough state variables have been measured (L is sufficient), estimating the unobserved states could be all one wishes to extract from data assimilation. In this

circumstance, examining the full state of the model may provide the understanding of the dynamics one requires (Wunsch and Heimbach 2007). As is clear from the examples in Chap. 2 and as will become clearer as we proceed, knowing how many measurements are required is a question one can answer with some confidence, but a bit of hard calculation. The issue is critical if the orbits of the dynamics are chaotic. The situation we have outlined is, however, plausible. We call it “qualifying” a model using data assimilation.

3.7 The Scientific Value of the Path Integral

As indicated earlier the path integral is an integral representation of the master equation for the stochastic process represented by the differential equation

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t), \mathbf{p}, \theta(t)) \quad (3.74)$$

in discrete time with $\theta(t)$, a stochastic element. Once the information transfer mechanism, namely, the measurement functions and their statistics, as well as the dynamics $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ and its statistics, is provided, we require reliable numerical methods for evaluating high-dimensional integrals. Approximations to these integrals in which the action is quadratic in the state variables represent linear problems and ignores the nonlinearities in the dynamics.

As an integral representation of the solution to a master equation, the path integral allows a global view of the solution going informatively beyond the local view provided by the PDE or methods that focus on constructing the probability distribution rather than physically interesting moments.

Approximations in which the action is not quadratic in \mathbf{X} but there appear only a finite number of moments around an expected path have a fundamental problem pointed out by Marcinkiewicz in 1939 and independently by Pawula in 1967a, b: the probability distributions will not be positive and the approximation is not self-consistent. This kind of approximation is at the heart of any closure procedure, such as the clever approach called “polynomial chaos.” This means one must have either linear dynamics or always consider all moments about the expected path as finite.

Familiar numerical evaluation methods for path integral, which we will discuss and illustrate in coming chapters, avoid the Marcinkiewicz/Pawula theorem (Marcinkiewicz 1939; Pawula 1967a,b). The path integral then represents a consistent format into which one may introduce approximations to the description of measurements and approximations for the nonlinear dynamics. It is the framework in which one can consistently provide quantification of the uncertainty in the estimation of model state variables, model parameters, and prediction using the statistical model. Furthermore, the formulation has no Gaussian constraints on the statistical properties entering the model.

It is interesting to note the role path integrals assume in statistical physics. [Chandler \(1987\)](#) notes in his introductory volume on statistical mechanics from 25 years ago that “Path Integral Monte Carlo is the method of choice for studying quantum systems at non-zero temperatures.”

3.8 Data Assimilation Path Integrals in Continuous Time

The path integrals considered until now have all been for discrete time. There is no essential difficulty in porting the results, in a formal sense, to flows in continuous time. The flow of the dynamics from a measurement time to the next measurement time is what we need to attend to. Measurements are still made discrete times in an observation window $[t_0, t_m = T]$. If one wishes to evaluate the path integral for utilization in a particular problem, with few exceptions, one will return to discrete time in any case.

We recall that if one wishes to refine the transport of the dynamics from one time to another through time steps much smaller than the time between measurements, one can refine the grid in time for that solution to the underlying differential equations as finely as one wishes. So the time steps for the dynamics can be decoupled from the time steps for the measurements in a natural fashion. The job then is to make the propagation of the dynamics from one measurement to the next flow in continuous time while holding fixed the finite number of measurements one makes in the interval $[t_0, T]$. To emphasize the difference between measurement times and flows of the dynamics, let’s call the measurement times $\{\tau_0, \tau_1, \dots, \tau_K\}$; these lie in the observation window $t_0 \leq \tau_0 < \tau_1 < \tau_2 \dots < \tau_K \leq T$.

There are two main issues we need to address: (1) transforming the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ into a continuous time format and (2) transforming the information transfer term into continuous time for the state variable but retaining discrete time for the measurement.

The first is handled by the formal discussion of ([Hochberg et al. 1999](#); [Zinn-Justin 2002](#)) and many others, though it is restricted to additive model errors. Suppose we begin with a set of ordinary differential equations for the model state variables $x_a(t)$; $a = 1, 2, \dots, D$

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)), \quad (3.75)$$

and to this we add some noise term $Q_a(t)$ acting in the D -dimensional space of the states. The differential equation becomes

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)) + Q_a(t). \quad (3.76)$$

Suppose we solve the differential equation for a specific realization of $q_a(t)$. Think of this as drawing a value for q_a from a distribution $P_Q(\mathbf{Q})$; $\mathbf{Q} = \{Q_1, Q_2, \dots, Q_D\}$ at each time. Having made the draws, the specific realization is just a set of constants affecting the solution to the model differential equation yielding the trajectory $x_a^{\mathbf{Q}}(t)$. The mean value over many draws is

$$\langle x_a(t) \rangle = \int [D\mathbf{Q}] x_a^{\mathbf{Q}}(t) P_Q(\mathbf{Q}). \quad (3.77)$$

The integral is over the whole path of values of \mathbf{Q} over the interval of time considered. As $\mathbf{Q}(t)$ is a function of time realized as a continuous variable, this is an integral over functions. Generally, if we want the average of a function, $G(x_a^{\mathbf{Q}}(t))$, of the $x_a^{\mathbf{Q}}(t)$'s, we write the expected value as

$$\langle G(x_a^{\mathbf{Q}}(t)) \rangle = \int [D\mathbf{Q}] G(x_a^{\mathbf{Q}}(t)) P_Q(\mathbf{Q}). \quad (3.78)$$

The idea is to now transform this from a functional integral over $\mathbf{Q}(t)$, the vector of noise to an integral over the state $\mathbf{x}(t)$. The transformation from the $\mathbf{Q}(t)$ integral to the $\mathbf{x}(t)$ integral as usual involves the Jacobian:

$$D\mathbf{Q} = \frac{\partial(\mathbf{Q})}{\partial(\mathbf{x})} D\mathbf{x}, \quad (3.79)$$

and since $\mathbf{Q}(t) = \frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t))$, the Jacobian of the transformation is

$$\frac{\partial(Q_a(t))}{\partial(x_b(t'))} = \det \left[(\delta(t-t')) \left(\delta_{ab} \frac{d}{dt} - DF(\mathbf{x}(t))_{ab} \right) \right], \quad (3.80)$$

where

$$DF_{ab}(\mathbf{x}(t)) = \frac{\partial F_a(\mathbf{x}(t))}{\partial x_b(t)}. \quad (3.81)$$

The elements in this determinant or transformation Jacobian are operators on functions of $\mathbf{x}(t)$.

Using this variable change, we arrive at

$$\langle G(\mathbf{x}) \rangle = \int D\mathbf{x} \det \left[\mathbf{I} \frac{d}{dt} - DF(\mathbf{x}(t)) \right] G(\mathbf{x}) P_Q \left(\frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t)) \right). \quad (3.82)$$

Following the argument in [Hochberg et al. \(1999\)](#) and [Zinn-Justin \(2002\)](#) we can express the determinant as

$$\begin{aligned} & \exp \left[\frac{1}{2} \int dt \sum_{a=1}^D \frac{\partial F_a(\mathbf{x}(t))}{\partial x_a(t)} \right], \\ & = \exp \left(\frac{1}{2} \int dt \text{trace}[DF(\mathbf{x}(t))] \right) \end{aligned} \quad (3.83)$$

along with factors not involving the state variable. These additional factors cancel in the mean value

$$\langle G(\mathbf{x}) \rangle = \frac{\int \mathcal{D}\mathbf{x} \exp\left(\frac{1}{2} \int dt \text{trace}[DF(\mathbf{x}(t))]\right) G(\mathbf{x}) P_Q \left[\frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t)) \right]}{\int \mathcal{D}\mathbf{x} \exp\left(\frac{1}{2} \int dt \text{trace}[DF(\mathbf{x}(t))]\right) P_Q \left[\frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t)) \right]}. \quad (3.84)$$

The term $P_Q \left[\frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t)) \right]$ is the continuous time analog of the model error contribution to the action. As a reminder, it is given here only when there is additive model noise, and there may be a more general replacement to the discrete time transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ in other cases of stochastic elements in the model.

That addresses moving the system between measurement times. In such intervals there are no conditional mutual information factors

$$\log[P(\mathbf{y}(n)|\mathbf{x}(n), Y(n-1))], \quad (3.85)$$

in the action associated with measurements, because there are no measurements. This factor in the action appears only at times τ_n . We incorporate this by introducing a function very peaked near the measurement time, achieving a value unity close to and around each τ_n . If we call this function the measurement weighting function $MW(t - \tau_n)$, then an example is given by

$$MW(t - \tau_n) = \frac{1}{\cosh^2\left(\frac{(t-\tau_n)}{w_n}\right)}, \quad (3.86)$$

and surely there are many more. The width w_n indicates that physically, there is some time interval over which the measurement is actually made and allows formal derivatives of the action with respect to the path element $\mathbf{x}(t)$ in the neighborhood of τ_n to be defined. The term in the action

$$- \sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), Y(n-1))], \quad (3.87)$$

we now write as

$$- \int_{t_0}^T dt \sum_{k=0}^K MW(t - \tau_k) \log[P(\mathbf{y}(\tau_k)|\mathbf{x}(t), \mathbf{Y}(k-1))], \quad (3.88)$$

in which we now have $\mathbf{Y}(k) = \{\mathbf{y}(\tau_k), \mathbf{y}(\tau_{k-1}), \mathbf{y}(\tau_{k-2}), \dots, \mathbf{y}(\tau_0)\}$, namely, all measurements up to the time τ_k . This is the factor to be inserted into the action for the continuous time general expectation value represented by Eq. (3.84).

We can write the conditional expectation value for a function $G(\mathbf{x})$ on the path in the usual form

$$E[G(\mathbf{x})|\mathbf{Y}(K)] = \frac{\int \mathcal{D}\mathbf{x} G(\mathbf{x}) \exp[-A_0(\mathbf{x}, \mathbf{Y}(K))]}{\int \mathcal{D}\mathbf{x} \exp[-A_0(\mathbf{x}, \mathbf{Y}(K))]}, \quad (3.89)$$

where now

$$\begin{aligned} A_0(\mathbf{x}, \mathbf{Y}(K)) &= -\frac{1}{2} \int dt \text{trace}[DF(\mathbf{x}(t))] \\ &\quad - \int_{t_0}^T dt \sum_{k=0}^K \text{MW}(t - \tau_k) \log[P(\mathbf{y}(\tau_k)|\mathbf{x}(t), \mathbf{Y}(k-1))], \\ &\quad - \log \left[P_Q \left(\frac{d\mathbf{x}(t)}{dt} - \mathbf{F}(\mathbf{x}(t)) \right) \right]. \end{aligned} \quad (3.90)$$

The factor $\text{trace}[DF(\mathbf{x}(t))]$ is associated with the change in state space volume as one moves around guided by the dynamics ([Abarbanel 1996](#); [Kantz and Schreiber 2004](#))

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t)), \quad (3.91)$$

and over a long orbit traversing the system attractor, it is negative. For Hamiltonian systems it is zero. For most of the examples in this book it is zero or a negative constant. The exception is the Hodgkin–Huxley neuron equations.

3.9 Earlier Work on Path Integrals in Statistical Data Assimilation

The idea of a path integral for statistical physics problems certainly goes back to the work of [Osager and Machlup \(1953\)](#) some decades ago. A good review and survey of this work is found in [Hochberg et al. \(1999\)](#). The use of path integrals for the statistical data assimilation problem has many sources ([Pythian 1977](#); [Jouvet and Pythian 1979](#); [Alexander et al. 2005](#); [Apte et al. 2006](#); [Restrepo 2008](#)). Each of them analyzes the continuous time problem from the outset. This has ambiguities associated with the treatment of statistical quantities in the continuous time limit. Further, only the case of Gaussian model errors or measurement errors is part of the discussion.

Interestingly, in the analysis of statistical data assimilation using discrete time, going back to the works of [Cox \(1964\)](#) [Friedland and Bernstein \(1966\)](#) and [Jazwinski \(1970\)](#), the connection with path integrals was not made, and the

consideration of the information theoretic interpretation of a measurement was not drawn. The development of data assimilation when measurements at different times are not independent was also not considered. Finally, almost all analyses moved immediately to the use of Gaussian statistics for measurement, and model errors and consideration of measurements correlated between measurement times are uniformly absent.

3.10 Synopsis and Perspectives: Statistical Data Assimilation

The full answer to statistical data assimilation questions embodied in the conditional expectation value of a function $G(\mathbf{X})$ on the $\mathcal{D} = (m+1)D$ -dimensional model path $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(m)\}$, given possible L -dimensional measurements $\mathbf{Y}(m) = \{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(m)\}$ during the observation window

$$E[G(\mathbf{X})|\mathbf{Y}] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}, \quad (3.92)$$

gives a framework for many approximations which have been addressed over many years. This general formulation gives a context to the variational approximations of Laplace (1774) to the integral and gives a further setting for the evaluation of the full integral using Monte Carlo or, perhaps other approximation methods.

In its construction we learned about the transfer of information from the data $\{\mathbf{y}(n)\}$ to the model state $\mathbf{x}(n)$ through conditional mutual information quantities. This provides a method to evaluate the importance of this or that individual or group of measurements to the stability and accuracy of the overall integral.

An integral representation of the solution to a master equation for the conditional probability $P(\mathbf{x}(m)|\mathbf{Y}(m))$ is the classical analog to the much studied problem in quantum mechanics. There one has a rule for the effect of a measurement on the quantum development of the wave function from its value at the beginning of an observation window $\psi(\mathbf{r}, t_0)$, and we have the analog of “data assimilation without data” which uses the classical action as the core of the quantum propagator of the system between observations. While each of these, classical and quantum frameworks, provides a general context for approximations and for some broad statements about characteristics of solutions, until the classical Lagrangian in the quantum context and the deterministic equations of motion in the classical are specified, answers of interest about this dynamics or that remain to be investigated.

There is a viewpoint that emerges from the general formulation which is of some importance: the role of measurements is to act like a “forcing term” in the stochastic time development of a nonlinear system guiding it to the regions of state space where data tells it should occupy. The quantity which is transferred in this “forcing” to the uninformed orbit is information in the precise sense of Shannon (Fano 1961) through the conditional mutual information.

Chapter 4

Evaluating the Path Integral

Prologue to the Chapter

The path integral giving the integral representation of any physical question encountered in statistical data assimilation is addressed in this chapter. We first outline the three methods for evaluating the path integral. The first two are variants on a stationary path approximation to the integral and its corrections presented as an infinite series. The series is identical to perturbation theory corrections to statistical physics and field theory estimations of path integrals encountered in those analyses. We work solely in discrete time and discrete space, so the infinities of field theory do not appear.

In the chapter we give some formal “loop” approximations to the effective action for statistical data assimilation. To date these are not “useful” but they certainly provide an untread path for data assimilation questions.

In the formulation of the path integral, it is certainly necessary to have data to assimilate, a model into which to assimilate the information in the data. Furthermore, we must specify the connection between the observed data and the state variables of the model via a “measurement function”, and even further we must make some precise statement on how model errors are represented. Typically that is rather *ad hoc*, so we present a small sidebar on examining the self-consistency of that choice.

4.1 Guide to Methods for Estimating the Path Integral

In the previous chapter we have seen that in statistical data assimilation the quantities of physical interest are the expected values of functions $G(\mathbf{X})$ along the state space path $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)\}$ through an observation window or during a prediction epoch outside of measurement windows. In an observation

window with possible measurements at times $\{t_0, t_1, t_2, \dots, t_n, \dots, t_m = T\}$, these conditional expectation values have the form

$$E[G(\mathbf{X})|\mathbf{Y}(m)] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}]}, \quad (4.1)$$

with $\mathbf{Y}(m) = \{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(m)\}$. Terms involving $\mathbf{y}(n)$ alone cancel in this expectation value.

The “residual” action Eq. (3.61) $A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}$ providing the distribution of paths in $\mathcal{D} = (m + 1)D$ -dimensional space $\exp[-A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}}]$ is given as

$$\begin{aligned} A_0(\mathbf{X}, \mathbf{Y}(m))_{\mathbf{R}} &= - \sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))] \\ &\quad - \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \log[P(\mathbf{x}(0))]. \end{aligned} \quad (4.2)$$

When measurements are absent at any t_n , we remove the $\log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))]$ term for the missing measurement.

In the action, $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ produces the dynamical propagation from one measurement time t_n to the next t_{n+1} . For numerical stability reasons we may wish to propagate the dynamics through many intermediate time steps between t_n and t_{n+1} . If there are N_I such time steps of length Δ between the observation times $t_n = t_n + 0\Delta$ and $t_{n+1} = t_n + N_I\Delta: \{t_{n+\Delta}, t_{n+2\Delta}, \dots, t_{n+(N_I-1)\Delta}\}$, we use the Kolmogorov-Chapman relation to represent the propagation over the time $t_{n+1} - t_n$ segment as

$$\begin{aligned} P(\mathbf{x}(n+1)|\mathbf{x}(n)) &= P(\mathbf{x}(n+N_I\Delta)|\mathbf{x}(n+0\Delta)) \\ &= \int \prod_{k=0}^{N_I-1} d^D x(t_n + k\Delta) P(\mathbf{x}(n+(k+1)\Delta)|\mathbf{x}(n+k\Delta)), \end{aligned} \quad (4.3)$$

with $\Delta = (t_{n+1} - t_n)/N_I$. This adds N_I terms to the action of precisely the form discussed at the beginning of the previous Chapter when we considered data assimilation without data. An adaptive time-stepping routine might add this form of intermediate time step only at times where the dynamics has large variation in $\mathbf{x}(t)$.

How are we to perform the \mathcal{D} -dimensional integral involved in (4.94)? Analytic evaluation is possible in the uninteresting situation where the measurement errors are Gaussian and the dynamical model is linear. We will discuss three possible approaches:

- An expansion about a stationary path \mathbf{S}_0 , where

$$\frac{\partial A_0(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{X}} \Big|_{\mathbf{x}=\mathbf{s}_0} = 0, \quad (4.4)$$

when there are no model errors,

- An expansion about a stationary path, \mathbf{S}_0 , where

$$\frac{\partial A_0(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{X}} \Big|_{\mathbf{X}=\mathbf{S}_0} = 0, \quad (4.5)$$

when there are model errors

- A direct Monte Carlo evaluation based on approximating $\exp[-A_0(\mathbf{X}, \mathbf{Y})]$

4.2 Stationary Path Methods

Laplace's Method

The saddle point method of Laplace (1774; Debye 1909) for performing integrals is effectively utilized for integrals of the form

$$I(\alpha) = \int dz e^{-\alpha f(z)}, \quad (4.6)$$

as α becomes large.

In this limit, one approximates the integral by seeking a saddle point z_0 where $f'(z_0) = 0$; $f''(z_0) > 0$ and expanding the integral in the vicinity of z_0 . Near z_0 $f(z) \approx f(z_0) + f^{(2)}(z_0) \frac{(z-z_0)^2}{2} + f^{(3)}(z_0) \frac{(z-z_0)^3}{6} + f^{(4)}(z_0) \frac{(z-z_0)^4}{24} \dots$, and the integral may be approximated as

$$I(\alpha) \approx e^{-\alpha f(z_0)} \int_{-\infty}^{+\infty} dz e^{-\alpha [f^{(2)}(z_0) \frac{(z-z_0)^2}{2} + f^{(3)}(z_0) \frac{(z-z_0)^3}{6} + f^{(4)}(z_0) \frac{(z-z_0)^4}{24}]} \quad (4.7)$$

or

$$I(\alpha) \approx e^{-\alpha f(z_0)} \sqrt{\frac{2\pi}{\alpha f^{(2)}(z_0)}} \left\{ \int_{-\infty}^{\infty} \frac{du}{\sqrt{\pi}} e^{-u^2} e^{-A_3(z_0)u^3/\sqrt{\alpha}} e^{-A_4(z_0)u^4/\alpha} \right\}, \quad (4.8)$$

with $A_3(z_0) = \frac{f^{(3)}(z_0)}{6} \left(\frac{2}{f^{(2)}(z_0)} \right)^{3/2}$ and $A_4(z_0) = \frac{f^{(4)}(z_0)}{6f^{(2)}(z_0)^2}$, after changing variables to $u = \sqrt{\frac{\alpha f^{(2)}(z_0)}{2}} (z - z_0)$.

If there are many saddle points, this approximation is performed for each.

Expanding in α leads to

$$I(\alpha) \approx e^{-\alpha f(z_0)} \sqrt{\frac{2\pi}{\alpha f^{(2)}(z_0)}} \left\{ 1 + \frac{1}{8\alpha} \left(\frac{5f^{(3)}(z_0)^2}{3f^{(2)}(z_0)^3} - \frac{f^{(4)}(z_0)}{f^{(2)}(z_0)^2} \right) + O\left(\frac{1}{\alpha^2}\right) \right\}. \quad (4.9)$$

Path Integrals

We have integrals of this general format in the conditional expectation value of functions $G(\mathbf{X})$ of the path \mathbf{X}

$$E[G(\mathbf{X})|\mathbf{Y}(m)] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}, \quad (4.10)$$

and in the Laplace-like approximation to the integral, we seek a path \mathbf{S}_0 along which

$$\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{X}=\mathbf{S}_0} = 0. \quad (4.11)$$

The path is a vector in $\mathcal{D} = (m + 1)D$ -dimensional space. We drop the explicit dependence on the measurements \mathbf{Y} as they play only a passive role in the integration. $\alpha = 1$ here, and we do not have an expansion parameter with which to mimic Laplace's expression, yet.

The extremum of the action defines a saddle path along which we might expect the important contribution to the integral to be achieved. As with Laplace's method for the simpler integral (4.6), there are corrections to the term evaluating the integrand along \mathbf{S}_0 . We return to these corrections.

The statement of the extremum of the action along $\mathbf{X} = \mathbf{S}_0$ is a standard unconstrained numerical optimization problem where $\mathbf{A}_0(\mathbf{X})$ is the objective function in the space of \mathcal{D} state variables and NP parameters. There are many publicly available numerical optimization routines. We do not intend to review the numerous methods. That would be a monograph or two in itself.

We have found and extensively used two publicly available numerical optimization programs, SNOPT and IPOPT. SNOPT (Gill and Wright 1982) (Barclay et al. 1998) (Gill et al. 2002) (Gill et al. 2005, 1998) and IPOPT (Wächter and Biegler 2006) are well adapted to the kind of variational problem posed by (4.11). Each of these uses the direct method which varies all the parameters \mathbf{p} and all the state variables at each time in the observation window $\{t_0, t_1, \dots, t_m = T\}$, $\mathbf{X} = \{x_a(0), x_a(1), \dots, x_a(m)\}$; $a = 1, 2, \dots, D$, subject to any equality or inequality constraints present. The difference between the programs lies in the details of the search methods they utilize in path space. These differences are discussed in the documentation available for each program.

The advantage of these two methods relative to the many others in the vast literature on numerical optimization is that the optimization procedure is done in a sparse space, making various sparse matrix routines quite useful. Further when the numerical optimization routine has completed according to its internal rules and according to the accuracy dictated by the user, the full path \mathbf{X} through the observation window is reported. This enables us to look at the behavior of the unobserved state variables as functions of time, and there may often be interesting physical information revealed about the operation of the model as it is informed by the data.

We have determined that both of these optimization algorithms spend a large fraction of their time performing linear algebra routines locally in state and parameter space, and the availability of parallelizable linear algebra routines (Curtis et al. 2010) offers real opportunities for acceleration of the searches involved.

The work required to translate the differential equations of the dynamics along with the bounds on each state variable and each parameter is found in a series of Python scripts written by Bryan Toth (2010). This reduces the creation of a C++ program to solve the variational problem from weeks or more to a few minutes.

4.2.1 No Model Error

If there is no error in the model $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$; $a = 1, 2, \dots, D$, then each transition matrix in the Markov dynamics is a delta function

$$P(\mathbf{x}(n+1)|\mathbf{x}(n)) = \delta^D(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})). \quad (4.12)$$

This means that in performing the path integral we have mD equality constraints $\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$. The search over \mathbf{X} and \mathbf{p} in the saddle path requirement

$$\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{x}=\mathbf{s}} = 0, \quad (4.13)$$

which is $\mathcal{D} + NP$ dimensional for NP parameters \mathbf{p} , formally involves determining only $\mathbf{x}(0)$ and \mathbf{p} and is $D + NP$ dimensional when the constraints are imposed.

The path integral now reduces to an integration over $P(\mathbf{x}(0))$ and factors in the action associated with each measurement

$$A_0(\mathbf{X}, \mathbf{Y}(m)) = - \sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))] - \log[P(\mathbf{x}(0))], \quad (4.14)$$

and $\mathbf{x}(n)$ comes from iterating the dynamical rule starting from $\mathbf{x}(0)$: $\mathbf{g}(\mathbf{x}(1), \mathbf{x}(0), \mathbf{p}) = 0$ gives $\mathbf{x}(1)$ in terms of $\mathbf{x}(0)$ and \mathbf{p} ; $\mathbf{g}(\mathbf{x}(2), \mathbf{x}(1), \mathbf{p}) = 0$ gives $\mathbf{x}(2)$ in terms of $\mathbf{x}(0)$ and \mathbf{p} ; \dots ; $\mathbf{g}(\mathbf{x}(m), \mathbf{x}(m-1), \mathbf{p}) = 0$ gives $\mathbf{x}(m)$ in terms of $\mathbf{x}(0)$ and \mathbf{p} .

The saddle path method for this case is the minimization of $A_0(\mathbf{X}, \mathbf{Y}(m))$. If all the iterations of the dynamics $\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$ are carried out, then $A_0(\mathbf{X})$ is an extraordinarily complex function of $\mathbf{x}(0)$, \mathbf{p} for any nonlinear system, and the derivatives required to find \mathbf{S}

$$\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{x}=\mathbf{s}} = 0 \quad (4.15)$$

are formidable. The `direct` method keeps the functional dependence on each component of \mathbf{X} and imposes $\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$ as equality constraints. This is the working mode of SNOPT and IPOPT, and it is the method we have used in our examples.

When the measurements at each time t_n are independent and the measurement error is Gaussian, then $A_0(\mathbf{X}, \mathbf{Y}(m))$ Eq. (4.14) is proportional to

$$-\sum_{n=0}^m \left\{ \sum_{l,k=1}^L (h_l(\mathbf{x}(n)) - y_l(n)) \frac{[R_m(n)]_{lk}}{2} (h_k(\mathbf{x}(n)) - y_k(n)) \right\} - \log[P(\mathbf{x}(0))], \quad (4.16)$$

and we have a constrained least-squares minimization problem. We have allowed here for the measurements to be a known function of the model state variables: $h_l(\mathbf{x}(t_n))$. The reader will recognize this formulation of data assimilation as the one we used in Chap. 2 in all the examples we discussed there.

In the geophysical literature (Evensen 2009; Lorenc and Payne 2007; Kalnay 2003) this formulation of statistical data assimilation is known as 4DVar, specifically as strong 4DVar, indicating it is a variational principle with strong equality constraints. The method itself may be as old as Newton, and it appears in numerous scientific inquiries across vast bodies of literature.

While it is impossible to quote all papers that use this method to address data assimilation, we give a few recent examples in neurobiology (Huys et al. 2006; Huys and Paninski 2009; Abarbanel et al. 2011), toxicology (Lyons et al. 2008), systems biology (Rodriguez-Fernandez et al. 2006), cell biology (Beard 2005) (Panning et al. 2008), biomedical engineering (Horvath and Manini 2008), chemical engineering (Xiong et al. 2006), wastewater treatment (Mueller et al. 2002), biochemistry (Dochain 2003), coastal and estuarine modeling (Yang and Hamrick 2005), immunology (Swameye et al. 2003). This hardly represents the tip of the iceberg in the use of the stationary path method in nonlinear problems.

The methods for solving such problems are also quite diverse, and while a discussion and comparison of them is the subject of another monograph, there is one classical method which we note (Kirk 1970). The idea is to enforce the equality constraints of the dynamics using Lagrange multipliers. The multipliers become time-dependent dynamical variables, called dual, and the combination of the original state variables and the multipliers forms an overall Hamiltonian dynamical system which may be solved by various methods. For dissipative nonlinear systems, especially those with chaotic orbits, the numerical procedure can be very sensitive as the multiplier (dual) variables must be solved backward in time, and the negative Lyapunov exponents, harmless in forward time integrations, play a destabilizing role.

As indicated in Chap. 2 in the examples involving the Colpitts oscillator and the Malkus waterwheel, the problem stated as (4.15) can still be ill posed numerically if there are chaotic orbits in the dynamics $\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$. We return to this matter in a later section.

4.2.2 Model Errors

If there are errors in our model differential equations

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t), \mathbf{p}) \quad (4.17)$$

or our discrete time maps coming from these differential equations

$$\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0, \text{ or } \mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p}) = 0, \quad (4.18)$$

then we require terms in the action to replace the delta functions.

The delta functions in the dynamics represent infinitely accurate resolution in the state space $\mathbf{x}(t)$. One form of model error is associated with recognizing we always have finite resolution and broadening the delta function in some manner. If we do that using the representation of the delta function as a Gaussian,

$$\delta^D(\mathbf{z}) = \sqrt{\frac{\det \mathbf{R}_f}{(2\pi)^D}} \exp\left[-\mathbf{z} \cdot \frac{\mathbf{R}_f}{2} \cdot \mathbf{z}\right], \quad (4.19)$$

in the limit that \mathbf{R}_f goes to infinity. Choosing \mathbf{R}_f large means taking the state space resolution to be of order $1/\sqrt{\det \mathbf{R}_f}$.

With this choice, the dynamical terms in the action take the form

$$\begin{aligned} & \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] = \\ & - \sum_{n=0}^{m-1} \left\{ \sum_{a,b=1}^D g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \frac{R_f(n)_{a,b}}{2} g_b(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \right\} \\ & + \text{constants.} \end{aligned} \quad (4.20)$$

The constants cancel in the conditional expectation values $E[G(\mathbf{X})|Y(m)]$. The $D \times D$ matrix $\mathbf{R}_f(n)$ contains the information on the broadening of the delta function for deterministic dynamics, this particular form for the broadening has independent model errors at each time t_n .

There are many, many ways to represent model errors associated with environmental noise or loss of resolution. Gaussian broadening is not special, though a common practice. Representing model errors associated with missing terms in the dynamical equations does not have any universally useful expression.

In the same geophysical literature (Evensen 2009; Lorenc and Payne 2007; Kalnay 2003) as above, this expression of model error is known as “weak 4DVar”

as the influence of the dynamics is spread over regions of phase space inversely proportional to the square root of the eigenvalues of the matrix $\mathbf{R}_f(n)$. In a sense strong 4DVar comes from weak 4DVar as “ $\mathbf{R}_f'' \rightarrow \infty$ ”.

The import of the previous two sections is to demonstrate that the familiar variational principles used widely across fields for the estimation of states and parameters are approximations to a common formulation: the path integral. Within the variational formulations, one derives a single path in state space, we have called it \mathbf{S}_0 , and no variations about that path are addressed. We should, as in the original Laplace method on which the saddle path methods were formulated, be able to evaluate, usually numerically and approximately, the corrections to the “leading” order we have displayed. Not only will there be estimates of the variations about the trajectory \mathbf{S}_0 , but some estimate of the size of the corrections and the convergence of the series of approximations ought to be available. Those variations are, of course, fully represented in the path integral, and we will give methods in later sections to evaluate these, systematically from the path integral.

4.3 Beyond the Stationary Path: Loop Expansions

4.3.1 *The Effective Action: Numerical Optimization Reappears*

The approximation to the path integral by a stationary path captures the sense of a mean path through the observation window. In the discussion of Laplace’s method we saw that it should be seen as the beginning of an expansion in a series of approximations. Variations about this path are also present in the path integral, and we consider now one method for evaluating those variations as a collection of further approximations to the integral as a whole.

We proceed by recognizing that as ever with statistical problems, variations about a mean value are contained in the moments about that mean or expected value. We follow practice in field theory, statistical physics, and actually statistics generally formulated (Feller 1971; Zinn-Justin 2002) by starting with the generating function for the moments of a distribution. This involves introducing the generating function $C(\mathbf{K})$ for cumulants of the distribution in the form

$$\exp[C(\mathbf{K})] = E[e^{\mathbf{K}\cdot\mathbf{X}}|\mathbf{Y}(m)] = \frac{\int d\mathbf{X} e^{\mathbf{K}\cdot\mathbf{X}} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}, \quad (4.21)$$

in which $\mathbf{K} = \{K_\nu\} = \{k_a(0), k_a(1), \dots, k_a(m)\}$; $a = 1, 2, \dots, D$ is a \mathcal{D} -dimensional constant vector. The \mathcal{D} -dimensional index $\nu = \{a, n\}$ stands for the vector index a along with the time index n needed to label a full path $\mathbf{X} = \{x_a(0), x_a(1), \dots, x_a(m)\}$. With this definition $C(\mathbf{K} = 0) = 0$.

The first derivative of $C(\mathbf{K})$ is

$$\begin{aligned} \frac{\partial C(\mathbf{K})}{\partial K_\nu} &= S_\nu(\mathbf{K}) \\ &= \frac{\int d\mathbf{X} X_\nu e^{\mathbf{K}\cdot\mathbf{X}} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} e^{\mathbf{K}\cdot\mathbf{X}} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}, \end{aligned} \quad (4.22)$$

and we see that

$$S_\nu(\mathbf{K} = 0) = \frac{\int d\mathbf{X} X_\nu \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]} = E[X_\nu | \mathbf{Y}(m)] \quad (4.23)$$

is the expected value of the path through the measurement window.

The second derivative of $C(\mathbf{K})$ evaluated at $\mathbf{K} = 0$ gives us the second moment about \mathbf{S}

$$\left. \frac{\partial^2 C(\mathbf{K})}{\partial K_\nu \partial K_\mu} \right|_{\mathbf{K}=0} = E[(X_\nu - S_\nu(\mathbf{K} = 0))(X_\mu - S_\mu(\mathbf{K} = 0)) | \mathbf{Y}(m)]. \quad (4.24)$$

Further derivatives yield higher moments about the mean value $S_\nu(K)$, namely, the cumulants (Feller 1971).

The reader may notice that we have now made a distinction between the path \mathbf{S}_0 where the action is an extremum and the expected value of the path \mathbf{S} . The former indicates the mode of the distribution $\exp[-A_0(\mathbf{X}, \mathbf{Y})]$, while the latter is the mean value.

4.3.2 The Effective Action

Taking hints from statistical physics, we are led via a Legendre transform of the cumulant generating function $C(\mathbf{K})$ to an effective action $A(\mathbf{S})$ by defining

$$A(\mathbf{S}) = -C(\mathbf{K}) + \sum_\nu K_\nu S_\nu(\mathbf{K}) = -C(\mathbf{K}) + \mathbf{K} \cdot \mathbf{S}(\mathbf{K}); \quad (4.25)$$

repeated indices are summed over. Evaluating the derivative of the effective action with respect to \mathbf{S} , we find

$$\begin{aligned} \frac{\partial A(\mathbf{S})}{\partial S_\nu(\mathbf{K})} &= -\frac{\partial K_\mu}{\partial S_\nu(\mathbf{K})} \frac{\partial C(\mathbf{K})}{\partial K_\mu} + \frac{\partial K_\mu}{\partial S_\nu(\mathbf{K})} S_\mu(\mathbf{K}) + K_\nu \\ \frac{\partial A(\mathbf{S})}{\partial S_\nu(\mathbf{K})} &= -\frac{\partial K_\mu}{\partial S_\nu(\mathbf{K})} S_\mu(\mathbf{K}) + \frac{\partial K_\mu}{\partial S_\nu(\mathbf{K})} S_\mu(\mathbf{K}) + K_\nu \\ \frac{\partial A(\mathbf{S})}{\partial S_\nu(\mathbf{K})} &= K_\nu, \end{aligned} \quad (4.26)$$

so at $\mathbf{K} = 0$, the expected value of the path $\langle X_\nu \rangle = E[\mathbf{X}_\nu | \mathbf{Y}(m)]$ is given by \mathbf{S} satisfying

$$\frac{\partial A(\mathbf{S})}{\partial S_\nu} = 0. \quad (4.27)$$

We see from this that the effective action contains all of the information about the complete version of the expected value of the orbit \mathbf{X} through the measurement window as well as the information about all the fluctuations about that expected value. The orbit S_ν is the generalization of the saddle path $(S_0)_\nu$ coming from the raw action $A_0(\mathbf{X})$

$$\frac{\partial A_0(\mathbf{X})}{\partial X_\nu} \Big|_{\mathbf{X}=\mathbf{s}_0} = 0. \quad (4.28)$$

The second derivative of $A(\mathbf{S})$ is given by

$$\frac{\partial^2 A(\mathbf{S})}{\partial S_\nu(\mathbf{K}) \partial S_\mu(\mathbf{K})} = \frac{\partial K_\nu}{\partial S_\mu(\mathbf{K})} = \left[\frac{\partial^2 C(\mathbf{K})}{\partial K_\nu \partial K_\mu} \right]^{-1}, \quad (4.29)$$

namely, the inverse of the second moment tensor of variation about the mean orbit.

If we are able to evaluate an interesting approximation to the effective action, then the determination of the expected orbit once again becomes a numerical optimization challenge, and all our experience and tools from the stationary path approximation are required again.

4.3.3 Dyson–Schwinger Equations for Statistical Data Assimilation

The cumulant generating function $C(\mathbf{K})$ defined by the $\mathcal{D} = (m+1)D$ -dimensional integral

$$\exp[C(\mathbf{K})] = \int d\mathbf{X} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \quad (4.30)$$

is invariant under translation of the integration variable $\mathbf{X} = \{X_\alpha\}$; $\alpha = \{a, n\}$; $a = 1, 2, \dots, D$; $n = 0, 1, \dots, m$ by a constant displacement ξ .

The translation of any function $f(\mathbf{X})$ by the constant ξ : $f(\mathbf{X}) \rightarrow f(\mathbf{X} + \xi)$ is accomplished by the operation $f(\mathbf{X} + \xi) = \exp[\xi \cdot \nabla_{\mathbf{X}}] f(\mathbf{X})$. Therefore, we have

$$\exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \rightarrow \exp[\xi \cdot \nabla_{\mathbf{X}}] \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}]. \quad (4.31)$$

When we integrate this over all \mathbf{X} , there results

$$\begin{aligned}
& \int d\mathbf{X} \exp[\xi \cdot \nabla_{\mathbf{X}}] \left\{ \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \right\} \\
&= \int d\mathbf{X} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \\
&+ \sum_{r=1} \int d\mathbf{X} \frac{\xi_{v_1} \cdots \xi_{v_r}}{r!} \frac{\partial}{\partial X_{v_1}} \cdots \frac{\partial}{\partial X_{v_r}} \left\{ \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \right\} \\
&= \exp[C(\mathbf{K})] + 0,
\end{aligned} \tag{4.32}$$

as all terms with $r \geq 1$ vanish by partial integration when $\exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}]$ vanishes rapidly enough as $\mathbf{X} \rightarrow \infty$, as it does here.

The $r = 1$ term in this expression reads

$$\begin{aligned}
& \int d\mathbf{X} \frac{\partial \left(\exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \right)}{\partial X_v} = 0 \\
&= \int d\mathbf{X} \left(-\frac{\partial A_0(\mathbf{X})}{\partial X_v} + K_v \right) \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \\
&= \exp[C(\mathbf{K})] \left\{ \left\langle -\frac{\partial A_0(\mathbf{X})}{\partial X_v} \right\rangle + K_v \right\} = 0.
\end{aligned} \tag{4.33}$$

or

$$\left\langle \frac{\partial A_0(\mathbf{X})}{\partial X_v} \right\rangle = K_v = \frac{\partial A(\mathbf{S})}{\partial S_v}, \tag{4.34}$$

which are the ‘‘classical’’ (deterministic) equations of motion in discrete time but including all statistical fluctuations associated with measurement errors, model errors, and initial state uncertainty. These are also the statistical physics equivalent of the Ehrenfest equations familiar from quantum theory. The additional Schwinger–Dyson equations (Zinn-Justin 2002) come from derivatives of 4.33 with respect to S_α .

If the dynamical equations are quadratic in the state variables, then when the model errors are Gaussian, we have at most quartic terms in \mathbf{X} appearing in $A_0(\mathbf{X})$:

$$A_0(\mathbf{X}) = \sum_{r=0}^4 \frac{a_{v_1, v_2, \dots, v_r}^{(r)}}{r!} X_{v_1} \dots X_{v_r}, \tag{4.35}$$

and

$$\frac{\partial A_0(\mathbf{X})}{\partial X_\mu} = \sum_{r=0}^3 \frac{a_{\mu, v_1, v_2, \dots, v_r}^{(r+1)}}{r!} X_{v_1} \dots X_{v_r}. \tag{4.36}$$

The Dyson–Schwinger equation corresponding to (4.33) is then

$$\begin{aligned}
 & \exp[C(\mathbf{K})] a_\mu^{(1)} + a_{\mu,\alpha}^{(2)} \int d\mathbf{X} X_\alpha \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \\
 & + \frac{a_{\mu,\alpha,\beta}^{(3)}}{2} \int d\mathbf{X} X_\alpha X_\beta \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \\
 & + \frac{a_{\mu,\alpha,\beta,\sigma}^{(4)}}{6} \int d\mathbf{X} X_\alpha X_\beta X_\sigma \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] \\
 & = \exp[C(\mathbf{K})] K_\mu,
 \end{aligned} \tag{4.37}$$

which can be expressed as

$$\begin{aligned}
 & a_\mu^{(1)} + a_{\mu,\alpha}^{(2)} \exp[-C(\mathbf{K})] \frac{\partial \exp[C(\mathbf{K})]}{\partial K_\alpha} + \frac{a_{\mu,\alpha,\beta}^{(3)}}{2} \exp[-C(\mathbf{K})] \frac{\partial^2 \exp[C(\mathbf{K})]}{\partial K_\alpha \partial K_\beta} \\
 & + \frac{a_{\mu,\alpha,\beta,\sigma}^{(4)}}{6} \exp[-C(\mathbf{K})] \frac{\partial^3 \exp[C(\mathbf{K})]}{\partial K_\alpha \partial K_\beta \partial K_\sigma} = K_\mu,
 \end{aligned} \tag{4.38}$$

The operator

$$\exp[-C(\mathbf{K})] \frac{\partial^n \exp[C(\mathbf{K})]}{\partial K_{v_1} \cdots \partial K_{v_n}}. \tag{4.39}$$

acting on a function $\psi(\mathbf{K})$ yields

$$\prod_{r=1}^n \left(S_{v_r} + \frac{\partial}{\partial K_{v_r}} \right) \psi(\mathbf{K}). \tag{4.40}$$

Now recalling that

$$\frac{\partial C(\mathbf{K})}{\partial K_\mu} = S_\mu, \tag{4.41}$$

and

$$\frac{\partial^2 C(\mathbf{K})}{\partial K_\alpha \partial K_\beta} = \left[\frac{\partial^2 A(\mathbf{S})}{\partial S_\alpha \partial S_\beta} \right]^{-1} + S_\alpha S_\beta, \tag{4.42}$$

along with

$$\begin{aligned} & \exp[-C(\mathbf{K})] \frac{\partial^3 \exp[C(\mathbf{K})]}{\partial K_\alpha \partial K_\beta \partial K_\sigma} \\ &= S_\sigma \Gamma(\mathbf{S})_{\alpha,\beta}^{-1} + S_\beta \Gamma(\mathbf{S})_{\alpha,\sigma}^{-1} S_\alpha \Gamma(\mathbf{S})_{\sigma,\beta}^{-1} + S_\alpha S_\beta S_\sigma + \frac{\partial^3 C(\mathbf{K})}{\partial K_\alpha \partial K_\beta \partial K_\sigma} \end{aligned} \quad (4.43)$$

and

$$\frac{\partial^3 C(\mathbf{K})}{\partial K_\alpha \partial K_\beta \partial K_\sigma} = -\Gamma(\mathbf{S})_{\sigma,\epsilon}^{-1} \Gamma(\mathbf{S})_{\alpha,\nu}^{-1} \Gamma(\mathbf{S})_{\tau,\beta}^{-1} \Gamma(\mathbf{S})_{\nu,\tau,\epsilon}^{(3)}. \quad (4.44)$$

We have written

$$\Gamma(\mathbf{S})_{\alpha,\beta} = \frac{\partial^2 A(\mathbf{S})}{\partial S_\alpha \partial S_\beta}$$

and

$$\Gamma^{(3)}(\mathbf{S})_{\alpha,\beta,\sigma} = \frac{\partial^3 A(\mathbf{S})}{\partial S_\alpha \partial S_\beta \partial S_\sigma}, \quad (4.45)$$

and this now gives an expression involving functions of \mathbf{S} alone.

There are an infinite number of constraints on the effective action $A(\mathbf{S})$ and its derivatives coming from the invariance of $C(\mathbf{K})$ under translations. The one usually discussed is linear in the action $A_0(\mathbf{X})$ (4.33); however, the remaining constraints, starting with

$$\int d\mathbf{X} \frac{\partial^2}{\partial X_\mu \partial X_\nu} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}] = 0, \quad (4.46)$$

is

$$\left\langle \frac{\partial A_0(\mathbf{X})}{\partial X_\alpha} \frac{\partial A_0(\mathbf{X})}{\partial X_\beta} - \frac{\partial^2 A_0(\mathbf{X})}{\partial X_\alpha \partial X_\beta} \right\rangle = K_\alpha K_\beta, \quad (4.47)$$

using

$$\left\langle \frac{\partial A_0(\mathbf{X})}{\partial X_\nu} \right\rangle = K_\nu, \quad (4.48)$$

are nonlinear in the action.

These relations among the “vertex functions”

$$\Gamma^{(n)}(\mathbf{S})_{\alpha_1, \dots, \alpha_n} = \frac{\partial^n A(\mathbf{S})}{\partial S_{\alpha_1} \cdots \partial S_{\alpha_n}}, \quad (4.49)$$

are “closure” relations one could derive from the original deterministic dynamics. If one is to use them for determining any of the $\Gamma(\mathbf{S})$, then it is important to avoid the negative probability consequences of the Marcinkiewicz and Pawula theorems (Marcinkiewicz 1939; Pawula 1967a,b).

There is another kind of Dyson–Schwinger equation which is of the form of a sum rule on the vertex functions, thus setting a scale for them. This is illustrated by the derivative of Eq. (4.34) with respect to the currents \mathbf{K} rather than the familiar derivatives with respect to \mathbf{S} just explored.

The derivative of

$$K_\nu = \frac{\int d\mathbf{X} \frac{\partial A_0(\mathbf{X})}{\partial X_\nu} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}]}, \quad (4.50)$$

with respect to K_μ , gives us

$$\delta_{\nu,\mu} = \left\langle \frac{\partial A_0(\mathbf{X})}{\partial X_\nu} X_\mu \right\rangle - \left\langle \frac{\partial A_0(\mathbf{X})}{\partial X_\nu} \right\rangle S_\mu, \quad (4.51)$$

which, as noted, sets a scale for the various expectation values of functions along the path.

As a closing item for this complicated section, it should be noted that the main use of these general relations among vertices has been to establish the renormalization characteristics of the underlying statistical theory. We have not addressed such issues here. As we deal with discrete space and time, the difficulties arising when one goes to continuous space and time (Zinn-Justin 2002) are not present. Whether these Dyson–Schwinger relations have interesting content remains to be explored.

4.3.4 The Effective Action: Loop Expansion

The effective action is more than a pleasant analog to the original action. If we can find a systematic manner in which to approximate $A(\mathbf{S})$, then to whatever order in that approximation we know the effective action, we may find quite directly the full orbit and fluctuations about that complete orbit by solving a variational problem. In other words, $A(\mathbf{S})$ contains the full information contained in the path integral. We follow the lead of statistical physics and field theory again (Zinn-Justin 2002).

To this end we introduce a parameter λ which will count “loops” in the diagrammatic representation of the expansion of the effective action. Using the definition of $A(\mathbf{S}) = -C(\mathbf{K}) + \mathbf{K} \cdot \mathbf{S}$, we may write

$$\exp[-A(\mathbf{S}) + \mathbf{K} \cdot \mathbf{S}] = \frac{\int d\mathbf{X} \exp[-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X}]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X})]}, \quad (4.52)$$

suggestive of how $\{A_0(\mathbf{X}), \mathbf{X}\} \rightarrow \{A(\mathbf{S}), \mathbf{S}\}$ through the operation of the path integral.

Now we introduce the loop parameter λ through

$$\exp[(-A(\mathbf{S}) + \mathbf{K} \cdot \mathbf{S})/\lambda] = \frac{\int d\mathbf{X} \exp[(-A_0(\mathbf{X}) + \mathbf{K} \cdot \mathbf{X})/\lambda]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X})]/\lambda} \quad (4.53)$$

and write

$$A(\mathbf{S}) = A_0(\mathbf{S}) + \lambda B(\mathbf{S}), \quad (4.54)$$

leading to

$$\begin{aligned} \exp[-B(\mathbf{S})] &= \int d\mathbf{X} \exp[-(A_0(\mathbf{X}) - A_0(\mathbf{S}) - (\partial A_0(\mathbf{S})/\partial \mathbf{S}) \cdot (\mathbf{X} - \mathbf{S}))/\lambda \\ &\quad + (\partial B(\mathbf{S})/\partial \mathbf{S}) \cdot (\mathbf{X} - \mathbf{S})], \\ &= \int d\mathbf{X} \exp[-(A_0(\mathbf{X}) - A_0(\mathbf{S}) - (\partial A_0(\mathbf{S})/\partial \mathbf{S}) \cdot (\mathbf{X} - \mathbf{S}))/\lambda \\ &\quad + B'(\mathbf{S})(\mathbf{X} - \mathbf{S})], \end{aligned} \quad (4.55)$$

recalling that

$$\frac{\partial A(\mathbf{S})}{\partial \mathbf{S}_\nu(\mathbf{K})} = \mathbf{K}_\nu = \frac{\partial[A_0(\mathbf{S}) + \lambda B(\mathbf{S})]}{\partial \mathbf{S}_\nu(\mathbf{K})}, \quad (4.56)$$

and denoting

$$B'(\mathbf{S})_\nu = \frac{\partial B(\mathbf{S})}{\partial \mathbf{S}_\nu}. \quad (4.57)$$

Next we expand $(A_0(\mathbf{X}) - A_0(\mathbf{S}) - (\partial A_0(\mathbf{S})/\partial \mathbf{S}) \cdot (\mathbf{X} - \mathbf{S}))/\lambda$ in a Taylor series in $\mathbf{X} - \mathbf{S}$ finding

$$\begin{aligned} &(A_0(\mathbf{X}) - A_0(\mathbf{S}) - (\partial A_0(\mathbf{S})/\partial \mathbf{S}) \cdot (\mathbf{X} - \mathbf{S}))/\lambda \\ &= \frac{1}{2\lambda} (\mathbf{X} - \mathbf{S})_\nu \gamma_0(\mathbf{S})_{\nu,\mu} (\mathbf{X} - \mathbf{S})_\mu + \sum_{r=3} \frac{A_0^{(r)}(\mathbf{S})_{\nu_1, \nu_2, \dots, \nu_r}}{r! \lambda} (\mathbf{X} - \mathbf{S})^{(r)}, \end{aligned} \quad (4.58)$$

where we designate the r^{th} derivative of $A_0(\mathbf{S})$ as

$$\frac{\partial^r A_0(\mathbf{S})}{\partial \mathbf{S}_{\nu_1} \partial \mathbf{S}_{\nu_2} \dots \partial \mathbf{S}_{\nu_r}} = A_0^{(r)}(\mathbf{S})_{\nu_1, \nu_2, \dots, \nu_r}, \quad (4.59)$$

abbreviate

$$(\mathbf{X} - \mathbf{S})_{v_1} (\mathbf{X} - \mathbf{S})_{v_2} \dots (\mathbf{X} - \mathbf{S}_{v_r}) = (\mathbf{X} - \mathbf{S})^{(r)}, \quad (4.60)$$

and write the $\mathcal{D} \times \mathcal{D}$ matrix

$$\frac{\partial^2 A_0(\mathbf{S})}{\partial \mathbf{S}^2} = A^{(2)}(\mathbf{S}) = \gamma_0(\mathbf{S}). \quad (4.61)$$

This suggests the change of variables

$$\chi = \sqrt{\frac{\gamma_0(\mathbf{S})}{2\lambda}} (\mathbf{X} - \mathbf{S}) - \sqrt{\frac{2\lambda}{\gamma_0(\mathbf{S})}} B'(\mathbf{S}), \quad (4.62)$$

giving us the expression

$$\begin{aligned} \exp[-B(\mathbf{S})] &= \frac{e^{\frac{1}{2} B'(\mathbf{S}) \gamma_0(\mathbf{S})^{-1} B'(\mathbf{S})}}{(\det[\frac{\gamma_0(\mathbf{S})}{2\pi\lambda}])^{1/2}} \\ \int \frac{d\chi}{\pi^{\mathcal{D}/2}} e^{-\chi' \chi} \exp[-\sum_{r=3} \frac{2^{r/2} \lambda^{r/2-1} A_0^{(r)}(\mathbf{S})}{r!} (\gamma_0(\mathbf{S})^{-1/2} \chi + \sqrt{\lambda/2} \gamma_0(\mathbf{S})^{-1} B'(\mathbf{S}))^r]. \end{aligned} \quad (4.63)$$

Expressing $B(\mathbf{S})$ as $B(\mathbf{S}) = \sum_{k=0} \lambda^k B_k(\mathbf{S})$, we can expand this representation in λ .

The net result for the effective action to second order in λ is derived in detail by [Zinn-Justin \(2002\)](#), Appendix A7:

$$\begin{aligned} A(\mathbf{S}) &= A_0(S) + \frac{1}{2} \text{trace} [\log \gamma_0(\mathbf{S})] \\ &+ \frac{1}{8} \sum_{v_1, v_2, v_3, v_4} A_0^{(4)}(\mathbf{S})_{v_1 v_2 v_3 v_4} [\gamma_0(\mathbf{S})_{v_1 v_2}^{-1} \gamma_0(\mathbf{S})_{v_3 v_4}^{-1}] \\ &- \frac{1}{12} \sum_{v, \mu, \tau, v_1, \mu_1, \tau_1} A_0^{(3)}(\mathbf{S})_{v \mu \tau} [\gamma_0(\mathbf{S})_{v v_1}^{-1} \gamma_0(\mathbf{S})_{\mu \mu_1}^{-1} \gamma_0(\mathbf{S})_{\tau \tau_1}^{-1}] A_0^{(3)}(\mathbf{S})_{v_1 \mu_1 \tau_1}. \end{aligned} \quad (4.64)$$

This is represented graphically in Fig. 4.1. Except for the matrix indices and the combinatorial coefficients, this is the generalization of the expansion of the one-dimensional integral that was treated by Laplace's method at the beginning of this Chapter.

The convergence of such a series is not an easy matter, and we refer to [Zinn-Justin \(2002\)](#) for a discussion of this and other issues. One comment of a physical nature is in order as we close this section: if one is dealing with problems in fluid dynamics, say, weather prediction or climate dynamics, the vector field in the

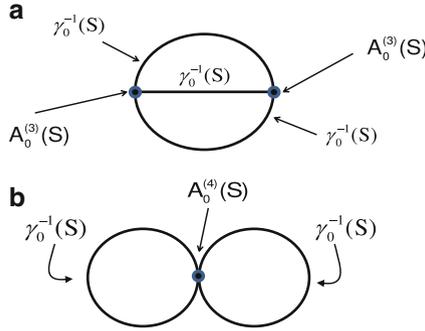


Fig. 4.1 The graphical representation of the order λ^2 contributions to the effective action. The elements in the graph are the derivatives of the action $A_0(\mathbf{S})$ with respect to \mathbf{S} , Eq. (4.59). The links in the graphs are the inverse of $A^{(2)}(\mathbf{S}) = \gamma_0(\mathbf{S})$ and acts as the “propagator” of information between nodes which are constructed using nonlinear terms in the action contributing to $r = 3$ and $r = 4$

nonlinear evolution equations $\mathbf{x}(n) \rightarrow \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ is often quadratic in components of the path \mathbf{X} , so only $A_0^{(3)}(\mathbf{S})$ and $A_0^{(4)}(\mathbf{S})$ will appear in the action, if the model error is Gaussian.

4.4 Estimating the Path Distribution $\exp[-A_0(\mathbf{X})]$

4.4.1 Langevin Equations: Fokker-Planck

Another view on evaluating integrals such as these recognizes that if we were able to estimate the distribution of paths \mathbf{X} and observations $\mathbf{Y}(m)$, $\exp[-A_0(\mathbf{X})]$, and identify N_{path} paths $\mathbf{X}^{(j)}$; $j = 1, 2, \dots, N_{path}$ drawn from that distribution, then we could use these to estimate conditional expectation values

$$E[G(\mathbf{X})|\mathbf{Y}(m)] \approx \frac{1}{N_{path}} \sum_{j=1}^{N_{path}} G(\mathbf{X}^{(j)}). \tag{4.65}$$

We can formally accomplish this by giving any path a “time” label s $\mathbf{X}(s)$ and an equation of motion in this time in the form of a Langevin equation, a stochastic differential equation, satisfying

$$\frac{dX_v(s)}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial X_v(s)} + \sqrt{2} \theta(s), \tag{4.66}$$

where $\theta(s)$ is Gaussian with mean zero and variance 1 and independent at each s .

A Langevin equation for paths $\mathbf{X}(s)$ that has Gaussian noise gives rise to a distribution of paths $P(\mathbf{X}, s)$ that satisfies a Fokker-Planck equation

$$\begin{aligned} \frac{\partial P(\mathbf{X}, s)}{\partial s} + \frac{\partial}{\partial \mathbf{X}_v} \left(-\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}_v} P(\mathbf{X}, s) \right) &= \frac{\partial^2 P(\mathbf{X}, s)}{\partial \mathbf{X}_v \partial \mathbf{X}_v}, \\ \frac{\partial P(\mathbf{X}, s)}{\partial s} &= \frac{\partial}{\partial \mathbf{X}_v} \left(P(\mathbf{X}, s) \left(\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}_v} + \frac{\partial \log P(\mathbf{X}, s)}{\partial \mathbf{X}_v} \right) \right) \end{aligned} \quad (4.67)$$

The s independent $P(\mathbf{X})$ satisfying this partial differential equation for large s is precisely $\exp[-A_0(\mathbf{X})]$, and this suggests that if we solve the Langevin equation for a set of N_{path} initial conditions $\mathbf{X}(s=0)$, remove the transients on each path starting from different $\mathbf{X}(s=0)$, and then we keep N_{path} paths for large s , we may approximate the expectation values as indicated in (4.65) using these N_{path} paths. There is no particular difficulty in this in principle; however, the gradient in \mathcal{D} -dimensional space $\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}_v}$ must be evaluated at each step in s , and the accuracy of numerical methods familiar for deterministic equations when adapted for stochastic differential equations is reduced (Burrage 1999; Burrage and Burrage 2002; Kloeden and Platen 1992). To achieve the same accuracy for the Langevin equation (4.66) as for deterministic versions ($\theta(s) = 0$) of the same equations thus requires substantially more effort.

The large s behavior of $P(\mathbf{X}, s)$ can also be analyzed in an elegant manner, as discussed by Klauder and Petersen (1985). If one writes

$$P(\mathbf{X}, s) = \exp[-A_0(\mathbf{X})/2] Q(\mathbf{X}, s), \quad (4.68)$$

then it can be verified that

$$\frac{\partial Q(\mathbf{X}, s)}{\partial s} = -\mathcal{H}Q(\mathbf{X}, s), \quad (4.69)$$

with the s -evolution operator \mathcal{H} given by

$$\begin{aligned} \mathcal{H} &= -\frac{\partial^2}{\partial \mathbf{X}^2} + \frac{1}{4} \left(\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}_v} \right)^2 - \frac{1}{2} \frac{\partial^2 A_0(\mathbf{X})}{\partial \mathbf{X}^2} \\ &= -\left(\frac{\partial}{\partial \mathbf{X}_v} - \frac{1}{2} \frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}_v} \right) \left(\frac{\partial}{\partial \mathbf{X}_v} + \frac{1}{2} \frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}_v} \right) \end{aligned} \quad (4.70)$$

The operator \mathcal{H} is Hermitian and positive semi-definite. The eigenfunction with eigenvalue 0 is $\exp[-A_0(\mathbf{X})/2]$, which is not a surprise. Variational principles for upper bounds for the next eigenvalue can be found in any textbook on quantum mechanics, and variational principles for lower bounds are in Bazley and Fox (1963); Sugar and Blankenbecler (1964). When the action increases as fast or faster than \mathbf{X}^2 , its spectrum is discrete.

The same issues, plus another one, arise in so-called Hybrid Monte Carlo methods where the Langevin equation is replaced by a Hamiltonian dynamics version where a pseudo-canonical momentum \mathbf{P} , conjugate to \mathbf{X} , is added to the problem, and orbits of the Hamiltonian

$$H(\mathbf{P}, \mathbf{X}) = \frac{\mathbf{P}^2}{2} + A_0(\mathbf{X}) \quad (4.71)$$

are evaluated. The canonical distribution of these orbits in $\{\mathbf{P}, \mathbf{X}\}$ space is

$$\exp[-H(\mathbf{P}, \mathbf{X})], \quad (4.72)$$

and reducing out the “momentum” part of the distribution yields $\exp[-A_0(\mathbf{X})]$. Unfortunately, Hamiltonian problems may have residual KAM tori and thus unwanted periodic orbits complicating the phase space distribution, and these need to be recognized and discarded by one device or another (Mackenzie 1989).

4.5 Monte Carlo Methods

Another approach to estimating the high-dimensional integrals is a direct evaluation using Monte Carlo methods. This seeks paths distributed as $\exp[-A_0(\mathbf{X})]$, and this has been explored in various forms for well over fifty years. We can hardly give the subject of Monte Carlo evaluation of integrals its full due in this book, but we’ll go over the essentials.

4.5.1 Metropolis–Hastings (Rosenbluth) Methods

The most widely used approach for selecting points in \mathbf{X} space with a specified distribution, here $\exp[-A_0(\mathbf{X})]$, is that pioneered by Rosenbluth and others in 1953 (Metropolis et al. 1953) and improved in Hastings (1970). It is one of a class of Monte Carlo methods that search through state space with biased random moves (Neal 1993; Rubin and Kroese 2008; Gamerman and Lopes 2006). The bias is the way the method directs random moves toward the desired distribution.

We start with a path $\mathbf{X}^{current}$ selected via a random choice or a more dynamical method. A sensible dynamical method in practice is to solve the variational principle

$$\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} = 0 \quad (4.73)$$

and use the approximate path, the mode of the distribution (Lorenz and Payne 2007) as a first selection of $\mathbf{X}^{current}$.

Once we have settled on a first path $\mathbf{X}^{current}$, we generate a new path \mathbf{X}^{new} from the current path $\mathbf{X}^{current}$ in two steps. First, a candidate path $\mathbf{X}^{proposed}$ is proposed by adding an unbiased random displacement to the current path $\mathbf{X}^{current}$. The displacement may be to any subset of the components of $\mathbf{X}^{current}$ and may be drawn from any distribution, as long as it is unbiased. This assures that the transition $\mathbf{X}^{current} \rightarrow \mathbf{X}^{proposed}$ is as likely as the transition $\mathbf{X}^{proposed} \rightarrow \mathbf{X}^{current}$. This implements a detailed balance between moving forward with this algorithm and moving in the opposite direction.

Next the proposed path is either accepted ($\mathbf{X}^{new} = \mathbf{X}^{proposed}$) or rejected ($\mathbf{X}^{new} = \mathbf{X}^{current}$) with a certain probability. The probability for acceptance is

$$P_{acceptance}(\mathbf{X}^{proposed}, \mathbf{X}^{current}) = \min(1, \exp[-\Delta A_0(\mathbf{X}^{proposed}, \mathbf{X}^{current})]), \quad (4.74)$$

where $\Delta A_0(\mathbf{X}^{proposed}, \mathbf{X}^{current}) = A_0(\mathbf{X}^{proposed}) - A_0(\mathbf{X}^{current})$ is the change in the action. This says that if a proposed change lowers the action, accept it. If the proposed change increases the action, it should be accepted with probability $\exp[-\Delta A_0(\mathbf{X}^{proposed}, \mathbf{X}^{current})]$. Only the change in action is required. Importantly, no derivatives of the action need be explicitly evaluated.

The accepted paths are collected into a set until we have N_{path} paths in our collection. Averages are now estimated using these paths.

4.5.2 Using GPU Parallel Processing

The Metropolis Monte Carlo method is simple and powerful, but it requires many path updates to achieve accurate statistics. One way to deal with this is to take advantage of parallel computing technology, using a graphics processing unit (GPU) (Quinn and Abarbanel 2011). With GPU technology it is possible to execute hundreds of threads of execution simultaneously on a single GPU. Typically each thread will perform the same operations, but on different pieces of the data. Since the paths are updated sequentially, the process cannot be run in parallel. Fortunately the many computations needed on each iteration can be done in parallel by having different threads work on different time steps. We will see some explicit examples of this computing advantage. A parallel implementation of the Metropolis–Hastings algorithm is sketched in Fig. 4.2.

First the current path $\mathbf{X}^{current}$ is set as an initial guess, and the observed time series \mathbf{Y} is loaded from a file. This current path includes the state vector $\mathbf{x}(n)$ at every time step $n = 0, 1, \dots, m$ as well as the NP parameters \mathbf{p} . The states and parameters are treated differently since the parameters are time independent, but the states vary in time. The current path, the observed time series, any external forcing, and a running sum of moments of the path components are allocated in GPU memory and initialized with the appropriate data.

The path update loop now begins. First the even time (n) states are updated followed by an update of the odd n states. This is done to uncouple the state vectors:

1. For an odd time step n , a thread does the following ($a = 1, 2, \dots, D; n = 0, 1, \dots, m$):
 - (a) Proposed change: $x_a(n)^{proposed} = x_a(n)^{current} + \Delta_a(n)U(-1, 1)$
 - (b) Calculate and store the vector field of the model equations: $\mathbf{F}(\mathbf{x}(n)^{proposed})$
 - (c) Evaluate the needed model error terms: $\mathbf{g}(\mathbf{x}(n)^{current}), \mathbf{g}(\mathbf{x}(n)^{proposed}),$
 $\mathbf{g}(\mathbf{x}(n+1)^{current}), \mathbf{g}(\mathbf{x}(n+1)^{proposed})$
 - (d) Calculate observation error terms: $\delta(\mathbf{x}(n)^{current}), \delta(\mathbf{x}(n)^{proposed})$
 - (e) Evaluate the change in the dynamical part of the action:
$$\Delta A_0(n)^d = \frac{R_f}{2} [(\mathbf{g}(\mathbf{x}(n)^{proposed})^2 - \mathbf{g}(\mathbf{x}(n)^{current})^2)]$$

$$+ [(\mathbf{g}(\mathbf{x}(n+1)^{proposed})^2 - \mathbf{g}(\mathbf{x}(n+1)^{current})^2)]$$
 - (f) Calculate the change in the observed data part of the action: $\Delta A_0(n)^o =$
$$\frac{R_m}{2} [(\delta(\mathbf{x}(n)^{proposed})^2 - \delta(\mathbf{x}(n)^{current})^2)]$$
 - (g) Evaluate the total change in the action: $\Delta A_0(n) = \Delta A_0(n)^d + \Delta A_0(n)^o$
 - (h) Use the acceptance probability: $P_{acceptance} = \min(1, e^{-\Delta A_0(n)})$ to accept or reject $\mathbf{x}(n)^{proposed}$.
 - (i) If $U(0, 1) < P_{acceptance}$ then accept the change: $\mathbf{x}(n)^{proposed} \rightarrow \mathbf{x}(n)^{new}$ and $\mathbf{F}(\mathbf{x}(n)^{proposed}) \rightarrow \mathbf{F}(\mathbf{x}(n)^{new})$.
 - (j) Otherwise reject $\mathbf{x}(n)^{proposed}$ and $\mathbf{x}(n)^{new} = \mathbf{x}(n)^{current}$.
2. For an even time step n , each thread executes the same steps shown above

Fig. 4.2 Flow of the parallel computing, Metropolis Monte Carlo state update. $U(a, b)$ is a random number between a and b drawn from a uniform distribution. $\delta(\mathbf{x}(n)) = \mathbf{y}(n) - \mathbf{x}(n)$

to calculate the change in action due to perturbing $\mathbf{x}(n)$, we need to know $\mathbf{x}(n-1)$ and $\mathbf{x}(n+1)$, but no other state vectors. This way each even n , and then each odd n , can be updated independently, in any order or simultaneously.

Once all the states are updated, then each parameter p_k is given a chance to change in sequence, $k = 1, 2, \dots, NP$. One parameter, p_k , is perturbed, and then $m+1$ threads are launched to calculate the change in the action. Each thread is assigned to one n , and it calculates the new vector field of the model differential equation $\mathbf{F}(\mathbf{x}(n)^{proposed})$. From this the total change in the action is calculated. Now a decision is made whether to accept the proposed change or reject it is made using the Metropolis rule. If the change is accepted, then $\mathbf{F}(\mathbf{x}(n)^{proposed}) \rightarrow \mathbf{F}(\mathbf{x}(n)^{new})$ for all n and $p_k^{proposed} \rightarrow p_k^{new}$.

After all the states and parameters have been updated, the current path can be used to update the path statistics. This is skipped for the first N_{init} path updates and after that only done every N_{skip} -th path update. The statistics collection happens on the GPU, also in parallel, and so the individual paths are not recorded. This avoids costly data transfers between the GPU and CPU.

4.5.3 Example Monte Carlo Problem: NaKL Neuron Model

We take the NaKL neuron model (2.41) as an example to illustrate the path integral Monte Carlo method (PIMC). This is a twin experiment differing from earlier discussions in that we now allow for both model error and measurement error. The use of the path integral will allow the estimation of expected values for the fixed parameters and the state variables as well as estimation of the RMS errors.

To perform the evaluation of the path integral we must specify how we represent the action. We will use the approximations discussed in Eqs. (4.16) and (4.20) in which we assume the noise in the measurements is Gaussian and the model errors are represented as a Gaussian broadening of the delta function $\delta^D(\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}))$ in the deterministic (no model error) case. Further assuming ignorance of any structure of the distribution of state variables at the beginning of observations $t_0 = 0$, we may write our approximation to the action as

$$A_0(\mathbf{X}, \mathbf{Y}(m)) = \frac{R_m}{2} \left\{ \sum_{n=0}^m \sum_{l=1}^L (y_l(t_n) - x_l(t_n))^2 \right\} + \left\{ \sum_{n=0}^{m-1} \sum_{a=1}^D \frac{R_f(a)}{2} g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) \right\}, \quad (4.75)$$

where $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})$ is the discrete form of the dynamics, R_m is proportional to the inverse of the noise level in the observations, and $R_f(a)$ is proportional to the inverse of the error level on the model state variable $x_a(t)$. We take the measurement errors and the model errors to be independent at different times.

Some Details For the Monte Carlo Evaluation

All of our Markov Chain Monte Carlo calculations worked with 10^7 sample paths; 10^3 were retained for calculations of moments and covariances of the distribution. The data assimilation window was comprised of $m + 1 = 4096$ points. The model error quantity $g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p})$ is a fourth-order Runge-Kutta integration scheme (Press et al. 2007) using a time step of $\Delta t = 0.01$ ms.

When the assimilation procedure was completed at the end of the observation window, $t = T$, a prediction using a fourth-order Runge-Kutta scheme was performed on each of the accepted paths using the state variables $\mathbf{x}(T)$ and the parameters associated with that path. Independent predictions were made for each accepted path. The predicted trajectories for $t > T$ were averaged to determine $\langle \mathbf{x}(t > T) \rangle$, and the RMS variation was evaluated about this mean. This gives the predicted quantities reported in the figures.

In order to assign values for R_m , the normalized deviation of the noise was estimated at 1 part in 10^3 for all dimensions of the model error. This normalized deviation was then scaled by the full range of the state variable and squared to get

the variance for that dimension, so for $V \in [-200, 200]$, $R_f = 6.25$, while for the gating variables, $R_f = 10^6$. We considered an experimental error of ± 1 mV giving $R_m \approx 1$.

We adjusted the size of the Monte Carlo step size using a scaling factor to achieve an acceptance rate near 0.5. The time required to perform each of our reported calculations with 10^7 candidate paths, each of dimension 16,402, took about an hour to complete on a single NVIDIA GTX 470 GPU. In our experience, provided that the dimension of the problem is roughly constant, the amount of time for a calculation scales roughly linearly with the number of CUDA cores on an NVIDIA GPU.

In practice, as the Metropolis–Hastings procedure seeks paths distributed about the maxima of the probability distribution $e^{-A_0(\mathbf{X})}$, a statistical minimization of $A_0(\mathbf{X})$ occurs when paths are accepted and rejected. This makes it a natural generalization of the variational procedure used in saddle path approximations but now applicable when there is model error as well. As emphasized, the data assimilation path integral approach also results in expected errors for estimations and predictions.

Another important feature of the data assimilation path integral approach is that no derivatives are required. This is a time-saving feature over the variational principles associated with saddle path calculations. In problems where the model may be specified with “switches” or step functions at thresholds, this is also a distinct advantage.

Parameter Estimation and RMS Errors in the Estimates

We selected a set of parameters as given in the “Value in ‘Data’” column in Table 4.1. The selection of maximal conductances g_{Na} , g_K , the values of g_L and E_L , along with the reversal potentials E_{Na} , E_K were taken from standard values found in many textbooks. The parameters in the kinetic equations for the gating variables $\{m(t), h(t), n(t)\}$ came from a fit to standard expressions for the time constants $\tau_i(V)$ and driving functions $a_{i0}(V)$ using the hyperbolic tangent functions in Eqs. (2.41)–(2.43) which are repeated for convenience:

$$\begin{aligned} \frac{dV(t)}{dt} &= \frac{1}{C} \left\{ g_{Na} m(t)^3 h(t) (E_{Na} - V(t)) + g_K n(t)^4 (E_K - V(t)) \right. \\ &\quad \left. + g_L (E_L - V(t)) + I_{DC} + I_{app}(t) \right\}, \\ \frac{da_i(t)}{dt} &= \frac{a_{i0}(V(t)) - a_i(t)}{\tau_i(V(t))} \quad a_{i0}(V) = \frac{1}{2} \left(1 + \tanh \left[\frac{(V - va)}{dva} \right] \right) \\ \tau_i(V) &= t_{a0} + t_{a1} \left(1 - \tanh^2 \left[\frac{(V - vat)}{dvat} \right] \right) \quad \text{or} \quad \tau_i(V) = t_{a0} + t_{a1} \tanh \left[\frac{(V - vat)}{dvat} \right], \end{aligned} \tag{4.76}$$

and $a = \{m, h, n\}$.

Table 4.1 NaKL twin experiment: path integral Monte Carlo estimation. Membrane voltage from an NaKL model with parameters specified in the “Value in ‘Data’ ” column was generated to create our twin experiment “data”

Name	Value in “data”	Estimated value	\pm RMS error
g_{Na}	120 mS/cm ²	113.443 mS/cm ²	3.33732 mS/cm ²
E_{Na}	50.0 mV	49.78762 mV	0.18938 mV
g_K	20 mS/cm ²	19.48006 mS/cm ²	0.7689 mS/cm ²
E_K	-77.0 mV	-77.23048 mV	0.2294 mV
g_L	0.3 mS/cm ²	0.27289 mS/cm ²	0.03132 mS/cm ²
E_L	-54.4 mV	-54.32392 mV	0.85092 mV
vm = vmt	-40.0 mV	-37.99831 mV	0.95009 mV
dvm = dvmt	15.0 mV	15.40739 mV	0.34118 mV
t_{m0}	0.1 ms	0.06095 ms	0.02714 ms
t_{m1}	0.4 ms	0.64496 ms	0.15212 ms
vh = vht	-60.0 mV	-61.73112 mV	0.59728 mV
dvh = dvht	-15.0 mV	-16.83959 mV	0.68188 mV
t_{h0}	1.0 ms	0.77615 ms	0.08225 ms
t_{h1}	7.0 ms	6.78265 ms	0.0463 ms
vn = vnt	-55.0 mV	-54.79809 mV	1.15982 mV
dvn = dvnt	30.0 mV	30.00895 mV	0.30394 mV
t_{n0}	1.0 ms	1.09374 ms	0.05155 ms
t_{n1}	5.0 ms	4.86783 ms	0.05321 ms

It is then presented to an NaKL model with unspecified parameters and unknown state variables. This is a “twin experiment” where the data and the model are known and the numerical experiment tests the methods used to pass information from observations to a model. The data assimilation path integral was used to provide the estimated parameter values, “estimated value”, and the RMS error in the estimation. Gaussian noise with RMS variation $\Delta V = \pm 1$ mV and zero mean was added to the voltage data presented to the NaKL model neuron

With some choice of initial conditions $\{V(0), m(0), h(0), n(0)\}$ we integrated the NaKL HH equations with an input current $I_{app}(t)$ consisting of a scaled waveform taken from the solution to a chaotic dynamical system. The amplitude of the waveform was selected so it stimulated the full dynamical range of the NaKL neuron activity. The solutions to the NaKL equations provide the “data” for the PIMC twin experiment. Gaussian noise of a level of ± 1 mV was added to the NaKL voltage output $V(t)$ to form the observations $y(t)$. With knowledge of the stimulating current, these voltages constitute our data.

To construct $A_0(\mathbf{X}, \mathbf{Y})$ we took $m+1 = 4096$ data points with $\Delta t = 0.01$ ms writing $\mathbf{g}(\mathbf{x}(n), \mathbf{x}(n+1), \mathbf{p}) = \mathbf{x}(n+1) - \Delta t \mathbf{f}(\mathbf{x}(n), \mathbf{p})$, where $\mathbf{f}(\mathbf{x}(n), \mathbf{p})$ is represented as an explicit fourth-order Runge-Kutta integration scheme. Using the methods described earlier, we evaluated expected values for the state variables $\langle x_a(n) \rangle$ and parameters through the observation period, and also evaluated second moments to yield RMS variations about these expected values. The dimension of the integral we are approximating is $4(4096) + 18 = 16,402$.

In Fig. 4.3 we display the estimated membrane voltage along with the estimated RMS error (blue dot \pm RMS error (red)) for the NaKL model between times $[0, 40.96]$ ms. The noisy data for the membrane voltage is shown in the solid black

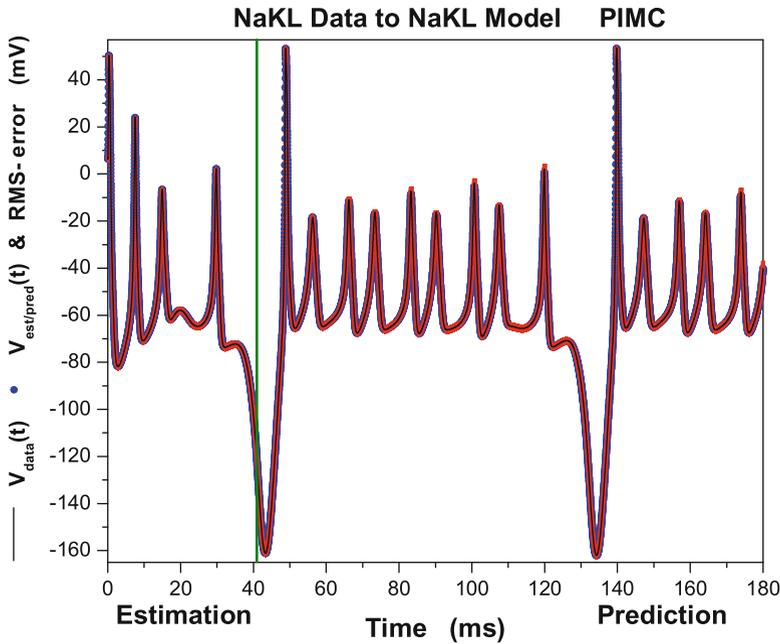


Fig. 4.3 Estimated and predicted membrane voltage when NaKL model data for the membrane voltage is presented to an NaKL model. The estimated membrane voltage in the observation window $[0, 40.96 = T]$ ms. V_{est} is a blue dot and the RMS errors are red. The known membrane voltage is shown as a black line. For $t \geq 41$ ms we display, with the same color scheme, the predicted membrane voltage using the estimated parameters and the estimated state variables at time T to initiate the predictions. The observation window and the prediction window are separated by a vertical green line at $t = 40.96$ ms. The observed voltage has a Gaussian error with zero mean and RMS error $\Delta V = \pm 1.0$ mV

line. For $t > 40.96$ ms we show the predicted value of the membrane voltage, again with RMS error as well as the known voltage data for $t > 41$ ms. In Fig. 4.4, we display the estimated and predicted values of the Na^+ activation variable, $m(t)$, an unobserved state variable, along with the known data for $m(t)$.

The accuracy with which the path integral estimates track the observed and unobserved states through the observation window is clear in these figures.

Bias in the Conditional Expected Values Arising from Model Error

While the errors are small in the estimates in Table 4.1 the expected or mean value appears to be biased away from the known value. This bias comes from our procedure and is associated with having model error as part of the action $A_0(\mathbf{X})$.

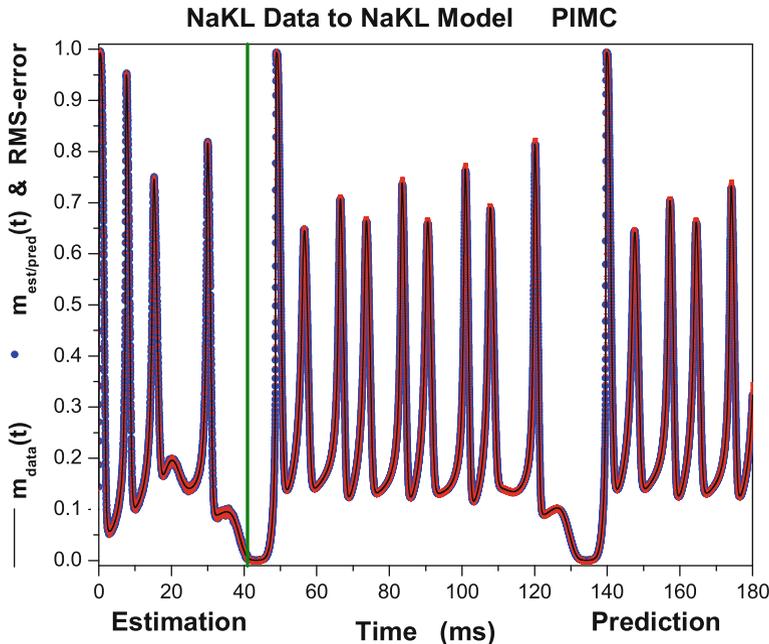


Fig. 4.4 Estimated and predicted Na^+ activation variable $m(t)$ when NaKL model data for the membrane voltage is presented to an NaKL model. The estimated $m(t)$ is shown in the observation window $[0, 40.96 = T]$ ms with a blue dot and the RMS errors with red. The known $m(t)$ is shown as a black line. For $t \geq 41$ ms we display, with the same color scheme, the predicted $m(t)$ using the estimated parameters and the estimated state variables at time T to initiate the predictions. The observation window and the prediction window are separated by a vertical green line at $t = 40.96$ ms. The observed voltage has a Gaussian error with zero mean and RMS error $\Delta V = \pm 1.0$ mV

The distribution of paths $\exp[-A_0(\mathbf{X})]$ is the solution to a Fokker-Planck equation of the form

$$\frac{d\mathbf{X}(s)}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}(s)} + \sqrt{2} N(0, 1), \quad (4.77)$$

where $\mathbf{X}(s)$ is the state space path as a function of “time” s and $N(0, 1)$ is Gaussian white noise with zero mean and variance unity. An equivalent to our Metropolis–Hastings Monte Carlo procedure is to solve this stochastic differential equation in $(m + 1)D$ -dimensions. There one can show that as $s \rightarrow \infty$, the distribution of paths is precisely $\exp[-A_0(\mathbf{X})]$. The Monte Carlo algorithm is seen as a search for extreme of the action along with accounting for the fluctuations about these extreme. All of this is guided by the observations as they enter the action.

To demonstrate the issue of biases in the estimation, suppose we had two measurements y_1, y_2 and two model outputs with the model taken as linear $x_2 = Mx_1$. Then the action we associate with this, including model error, is

$$A_0(x_1, x_2, y_1, y_2) = \frac{1}{2} [(y_1 - x_1)^2 + (y_2 - x_2)^2 + R(x_2 - Mx_1)^2], \quad (4.78)$$

and this has its minimum at

$$\begin{aligned} x_1 &= \frac{(1 + R)y_1 + RM y_2}{1 + R(1 + M^2)} \\ x_2 &= \frac{(1 + RM^2)y_2 + RM y_1}{1 + R(1 + M^2)}. \end{aligned} \quad (4.79)$$

This clearly shows the bias we anticipated. As $R \rightarrow \infty$, we see that the bias remains, but $x_2 = Mx_1$ is enforced. If $R = 0$, however, the minimum is at $x_1 = y_1$, $x_2 = y_2$, and if the dynamics underlying the data source satisfies $y_2 = My_1$, the same holds for the model.

Predicting Beyond the Observation Window

We now have a completed NaKL model as we have estimated the parameters of the model and evaluated RMS errors in those estimates. In Table 4.1 we present these estimates and the RMS errors of the biophysical parameters in the model. Using the estimated values of the parameters and of the state variables at T , we use the completed model to forecast the behavior of the model neuron for times $t > T$. We will have a spread in values for the state variables for $t > T$ as we have a spread of initial conditions at $t = T$. The predictions for the expected values are displayed with blue dots and the RMS error of the predictions with red lines. These are shown along with the known values displayed as black lines. In this twin experiment we are able to directly compare parameters and state estimates to the true values, but in biological experiments this is not possible.

In Fig. 4.5 we indicate how one of the parameters, normalized to unity using its known value in this twin experiment, behaves as a function of the number of Metropolis–Hastings Monte Carlo iteration steps used in evaluating the path integral. After some excursions while the Monte Carlo selection moves about through the space of all parameters and the state space of all state variables \mathbf{X} along the entire path, we achieve a statistically stationary distribution of values for this parameter. The expected value for this distribution should be 1.0, and the numerical estimates produce 1.02.

To address the practical questions of how long such calculations take to perform, precisely the same calculation was done over a range of time series lengths from $m = 100$ up to $m = 24,000$, with 1/10th as many iterations in the longer calculation. The execution time, T_{GPU} , was recorded. In all cases, the number of threads per block was set to 100. For comparison a similar, but not exactly the same, calculation was done on a single core of an Intel Core i3 CPU. A linear time scaling of $T_{CPU} = m \times (1.24 \text{ s})$ was seen from several trials on the CPU. In Fig. 4.6 the

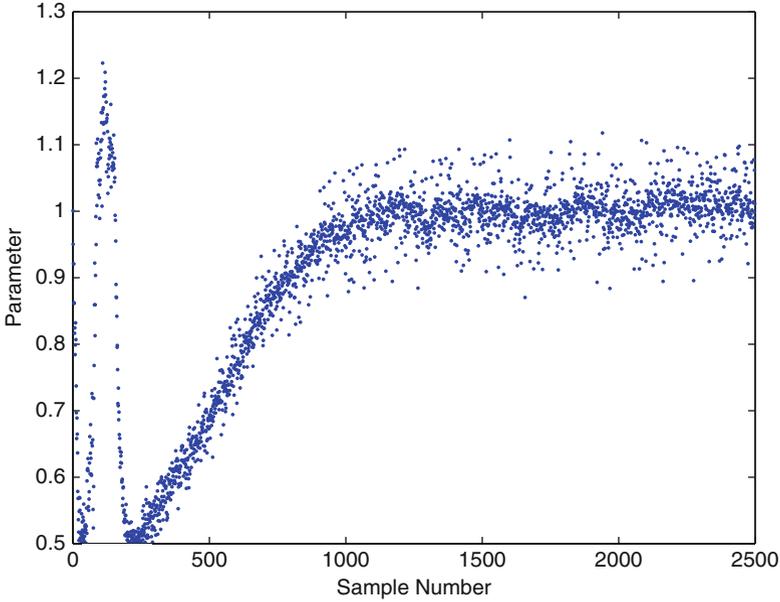


Fig. 4.5 The evolution of a NaKL parameter normalized to its known value as the PIMC path update procedure runs. The first 1,250 samples were discarded. Simulated annealing was applied during the first 250 iterations. Eventually the parameter fluctuates about a steady value having a mean of 1.02

parallel speedup factor, T_{CPU}/T_{GPU} , is displayed. When m is small, the full capability of the GPU is not utilized as the threads on the chip are underused. This means that larger problems will benefit from parallel evaluation on devices such as GPUs.

4.6 Consistency of Model Errors

If there are no model errors and the state space resolution of the state were perfect, the transition matrix in which the dynamics resides would be

$$P(\mathbf{x}(n+1)|\mathbf{x}(n)) = \delta^D(\mathbf{g}(\mathbf{x}(n), \mathbf{x}(n+1), \mathbf{p})), \quad (4.80)$$

where we denote the deterministic dynamics as $\mathbf{g}(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0$. If we can write this as an explicit expression $\mathbf{g}(\mathbf{x}(n), \mathbf{x}(n+1), \mathbf{p}) = \mathbf{x}(n+1) - \mathbf{x}(n) - \mathbf{f}(\mathbf{x}(n), p) = 0$. In this section where we perform Monte Carlo calculations of our path integral, we use

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \frac{1}{2} \left\{ \mathbf{F}(\mathbf{x}(n)) + \mathbf{F}(\mathbf{x}(n+1)) \right\}. \quad (4.81)$$

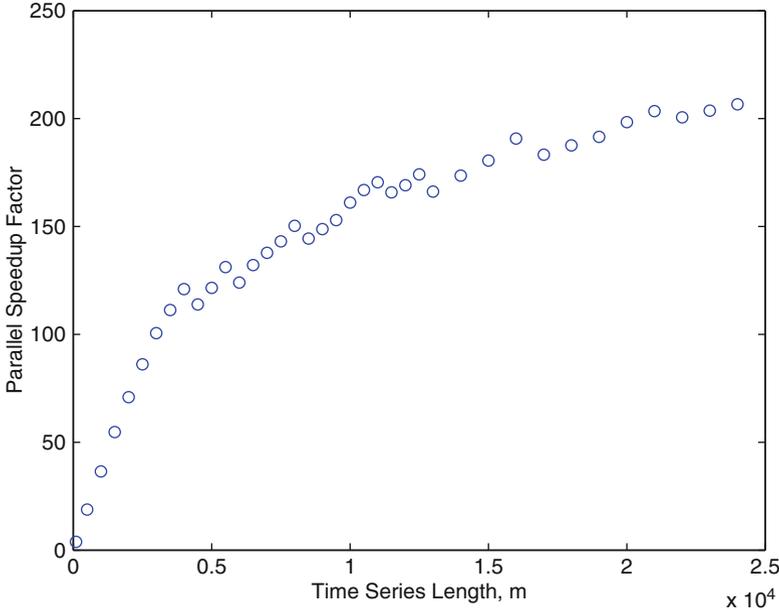


Fig. 4.6 The parallel speedup factor as a function of time series length m . The parallelspeed up factor is defined as T_{CPU}/T_{GPU} . The speedup increases as the problem size grows until the full resources of the GPU are being utilized. The maximum speedup appears to increase more or less linearly as the number of cores on the GPU is augmented (Quinn and Abarbanel 2011)

Perfect state space resolution is not possible, so we must make an approximation to $P(\mathbf{x}(n+1)|\mathbf{x}(n))$. If we assume the delta function is broadened by Gaussian noise or errors in resolution, we can replace

$$\delta^D(\mathbf{x}(n+1)-\mathbf{f}(\mathbf{x}(n), \mathbf{p})) \rightarrow \left(\frac{R_f}{2\pi}\right)^{D/2} \exp\left[-\frac{R_f}{2} \sum_{a=1}^D \sum_{n=0}^{m-1} (x_a(n+1)-f_a(\mathbf{x}(n), \mathbf{p}))^2\right], \tag{4.82}$$

which becomes the delta function as the inverse resolution $R_f \rightarrow \infty$. R_f is a scalar here. We recognize that its value might depend on the component a of the state variable $x_a(n)$, and it could connect model errors at different times. There are many approximate forms of a delta function which become the delta function when the width of the approximate distribution goes to zero, and the Gaussian case, though in common use, is here as an illustration. This section is devoted to looking at the self-consistency of this assumption about the distribution of model errors.

To use the general result for numerical estimation of $\langle G(\mathbf{X}) \rangle$, we also require an approximation for the conditional mutual information term in the action. Though not necessary, it is common to assume that the measurements $y_l(n)$ are independent of each other at time n and independent of measurements at earlier times. As an alert coming from the actual physical properties of measurement instruments, one

should look at Hamill (2006). In the case the measurement errors are independent at different observation times, we may represent the conditional mutual information terms in the action as

$$\frac{R_m}{2} \sum_{l=1}^L \sum_{n=0}^m (y_l(n) - h_l(\mathbf{x}(n)))^2, \quad (4.83)$$

when the measurement errors are assumed to be distributed as a Gaussian with a scalar variance R_m^{-1} . If the errors for various state variables are different, R_m is a matrix.

With these approximations, the action becomes

$$\begin{aligned} A_0(\mathbf{X}(m), \mathbf{Y}(m)) &= \frac{R_m}{2} \sum_{l=1}^L \sum_{n=0}^m (y_l(n) - h_l(\mathbf{x}(n)))^2 \\ &+ \frac{R_f}{2} \sum_{a=1}^D \sum_{n=0}^{m-1} (x_a(n+1) - f_a(\mathbf{x}(n), \mathbf{p}))^2 \\ &- \log[P(\mathbf{x}(0))]. \end{aligned} \quad (4.84)$$

When the dynamics $\mathbf{f}(\mathbf{x})$ is not linear, this is not quadratic in the state variables, so the integral for $E[G(\mathbf{X})|\mathbf{Y}(m)]$ requires numerical evaluation.

In this section we wish to examine the consistency of assumptions such as Eq. (4.82). If the Gaussian assumption in Eq. (4.82) or any other specific assumption about the stochastic model error is to be consistent, then the distribution of the model error for $a = 1, 2, \dots, D$; $n = 0, 1, \dots$

$$ME_a(n) = x_a(n+1) - f_a(\mathbf{x}(n)) \quad (4.85)$$

resulting from performing the integral with $G_{ME}(\mathbf{X}) = ME_a(n)$ should be the same as when evaluating the expectation of $G_{ME}(\mathbf{X})$ directly from the assumption on the transition matrix $P(\mathbf{x}(n+1)|\mathbf{x}(n))$.

This means that $\langle G_{ME}(\mathbf{X}) \rangle$ should numerically be the same as the expected value from the assumed model error distribution

$$\langle G_{ME}(\mathbf{X}) \rangle_{ME} = \int d\mathbf{X}(m) \prod_{n=0}^{m-1} P(\mathbf{x}(n+1)|\mathbf{x}(n)) G_{ME}(\mathbf{X}). \quad (4.86)$$

While demonstrating this would require comparing the quantities $\langle G_{ME}(\mathbf{X}) \rangle$ and $\langle G_{ME}(\mathbf{X}) \rangle_{ME}$ for arbitrary functions $G(\mathbf{X})$ or, perhaps, for all moments assuming they exist, we cannot do that in practice, so we examine a much more limited set of comparisons.

The reader will have realized that the equality of $\langle G_{ME}(\mathbf{X}) \rangle$ and $\langle G_{ME}(\mathbf{X}) \rangle_{ME}$ is equivalent to

$$\sum_{n=0}^m \text{CMI}(\mathbf{x}(n), \mathbf{y}(n) | \mathbf{Y}(n-1)) = 0 \quad (4.87)$$

and is essentially a test of how well the model, along with assumptions of representations of its stochastic errors, performs in matching the model output with the observations $\mathbf{Y}(m)$. This is also essentially the same as saying the model output $h_l(\mathbf{x}(n))$ at each measurement time t_n is synchronized to the measurements $y_l(n)$ at those times. In the case of independent measurements with a Gaussian distribution of measurement errors Eq. (4.82), it also represents maximal information transfer from the observed data to the model.

4.6.1 An Example from the Lorenz96 Model with $D = 100$

To examine an assumption about the distribution of model errors, we turn, once again, to the use of the Lorenz96 model. We look at two cases: one has $D = 100$ and the other utilizes $D = 20$.

The 1996 model of Lorenz (1996) has D degrees of freedom $x_a(t)$; $a = 1, 2, \dots, D$ in a periodic sequence satisfying the differential equations

$$\frac{dx_a(t)}{dt} = x_{a-1}(t)(x_{a+1}(t) - x_{a-2}(t)) - x_a(t) + f, \quad (4.88)$$

with $x_{-1}(t) = x_{D-1}(t)$, $x_0(t) = x_D(t)$, and $x_{D+1}(t) = x_1(t)$, with some choice of $x_a(0)$. The “forcing constant” f is a control parameter for the bifurcations of the solutions $x_a(t)$ of these equations, and when $f \approx 8$ or larger, the solutions exhibit chaotic behavior. We will encounter this model in the chapter on “twin experiments” where we explore different aspects of the model than addressed here.

We selected $D = 100$ and $f = 8.17$ and generated a set of “data” $y_l(t_n)$ for $t_{n+1} - t_n = 0.01$ and $L = 40$. $L = 40$ was chosen from considerations of the smoothness of the action for this problem as a function of the number of measurements or equivalently from the number of positive conditional Lyapunov exponents on the synchronization manifold (Abarbanel et al. 2010; Quinn and Abarbanel 2010). The measurement function was just $h_l(\mathbf{x}(n)) = x_l(n)$. We used the Gaussian assumption of stochastic noise to approximate the action in the data assimilation path integral, so we adopted Eq. (4.84) for the evaluation of $\langle G_{ME}(\mathbf{X}) \rangle$.

We generated “data” solving Eq. (4.88) for a time step $\Delta t = 0.01$ and evaluated $y_l(n)$; $l = 0, 1, \dots, D-1$ for $t_{n+1} - t_n = 0.05$, namely, every fifth time step in the development of the dynamics. Data for the 40 data time series

$$l = \{0, 2, 5, 7, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32, 35, 37, 40, 42, 45, \\ 47, 50, 52, 55, 57, 60, 62, 65, 67, 70, 72, 75, 77, 80, 82, 85, 87, 90, 92, 95, 97\}$$

were presented to the model for an observation window $0 \leq t \leq 4$. No observations were presented in a prediction window [4, 6]. The integrals involved were of dimension $(m + 1)D = 60100$, and we evaluated them approximately using a standard Metropolis–Hastings Monte Carlo method (Quinn and Abarbanel 2010; Neal 1993; Hastings 1970). In this implementation we selected starting paths at random and performed 35,000 initialization path selections before recording statistics on $\langle G_{ME}(\mathbf{X}) \rangle$ for another 310,000 paths. The recorded paths were broken into blocks of 100 with 3,100 blocks evaluated to approximate the integrals. These calculations were performed on a standard 2.5 GHz CPU.

In the calculations we averaged over each block of 100 paths to achieve $N_{AP} = 3100$ accepted paths $\mathbf{X}^{(j)}$ of length 60101 each including the parameter f . With these paths we approximated $\langle G_{ME}(\mathbf{X}) \rangle$ as

$$\langle G_{ME}(\mathbf{X}) \rangle \approx \frac{1}{N_{AP}} \sum_{j=1}^{N_{AP}} G_{ME}(\mathbf{X}^{(j)}). \quad (4.89)$$

By the Metropolis–Hastings construction the accepted paths were distributed according to $\exp[-A_0(\mathbf{X}, \mathbf{Y})]$. The expected error in this evaluation of $\langle G_{ME}(\mathbf{X}) \rangle$ is a few parts in 10^{-3} .

We evaluated the expected values of the set of functions $G(\mathbf{X}) \{x_a(n)^q, ME_a(n)\}$ for $q = 1, 2, 3, 4$ allowing us to estimate the expected mean path, the RMS variation about that path, as well as the skewness and kurtosis about that mean. The latter quantities allow us to examine whether the common assumption that the integrals involved are approximately Gaussian is correct. From the collection of accepted paths we are also able to estimate the marginal distributions of any element of the path, and in particular we were interested in the distribution of $ME_a(n)$. If the assumption made in formulating the action Equation (4.84) that the stochastic model errors are distributed as a Gaussian, then we expect the mean of $ME_a(n)$ for any index a and any time n to be zero. We expect the RMS error variation about this mean to satisfy $\text{RMS}(ME_a(n))\sqrt{R_f} = 1$ and the distribution $P(ME_a(n))$ for any a and n to be Gaussian.

When we generated our “data” by solving the Lorenz96 $D = 100$ equations, we added Gaussian noise with a signal-to-noise ratio of about 23 dB to the clean signal. This translates to $R_m \approx 8.0$, and we used that value in our Monte Carlo integrations. We also selected $R_f = 100$ as our experience with these methods (Quinn and Abarbanel 2010) suggests that $R_f \approx 10R_m$ gives a sufficiently large R_f that the imposition of the approximate equations of motion is accurate.

As examples of the outcome of these calculations, we report that for $ME_{76}(81)$ (chosen at random from among all the $ME_a(n)$), the mean value was 9.9×10^{-3} and $\text{RMS}(ME_{76}(81))\sqrt{R_f} = 0.97$, and the skewness and kurtosis of this variable were smaller than 0.01 in magnitude.

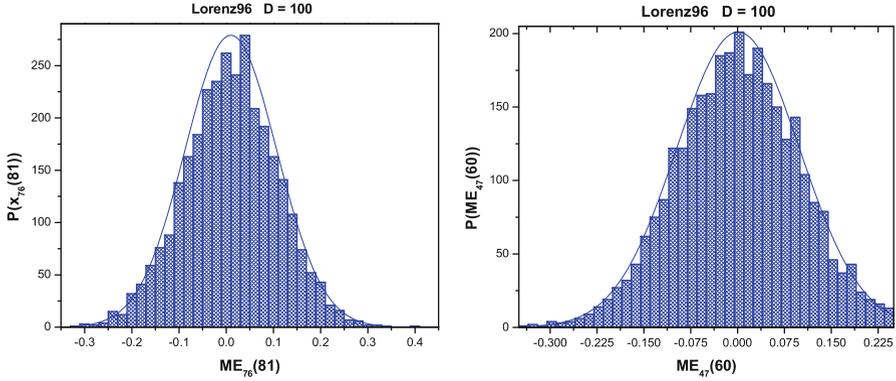


Fig. 4.7 *Left* Distribution of the stochastic model error $ME_{76}(81)$ from the $D = 100$ Lorenz96 model compared to a best-fit Gaussian distribution. *Right* Distribution of the stochastic model error $ME_{47}(60)$ from the $D = 100$ Lorenz96 model compared to a best-fit Gaussian distribution

Figure 4.7 shows the distribution of values of $ME_{76}(81)$ over the 3,100 block of paths along with the best-fit Gaussian to that distribution. As $RMS(ME_{76}(81)) \sqrt{R_f} = 0.97$, it is clear to the eye that this distribution of this model error is consistent with the assumed stochastic model error distribution. Figure 4.7 also shows the distribution of $ME_{47}(60)$ from the same set of calculations compared again to the best-fit Gaussian distribution. In this case the mean value of $ME_{47}(60)$ was 8.8×10^{-4} and $RMS(ME_{47}(60)) \sqrt{R_f} = 0.9899$; again the skewness and kurtosis are quite small as would be expected for a nearly Gaussian distribution. We can say for this case also that the output distribution of stochastic model error is consistent with the assumed distribution.

In Fig. 4.8 *Left* we show the expected value of the observed variable $x_{47}(t)$ and its RMS error through the observation window $[0, 4]$ and into the prediction region $[4, 6]$. In Fig. 4.8 *Right* we show the expected value of the unobserved variable $x_{76}(t)$ and its RMS error through the observation window $[0, 4]$ and into the prediction region $[4, 6]$. Figure 4.9 shows the skewness and kurtosis for the observed variable $x_{47}(t)$ through the window $[0, 6]$. We see that the skewness and kurtosis are small in the observation window and then grow substantially after observations are terminated. This is consistent with the chaotic orbits of the model. Figure 4.10 displays the skewness and kurtosis of the unobserved variable $x_{76}(t)$, and we see that the values both within and without the observation window are larger than for the observed variable.

We noted above that the consistency of the assumed and the calculated distribution is equivalent to the precision of the equality of model output as estimated through the path integral and the observations presented to the model.

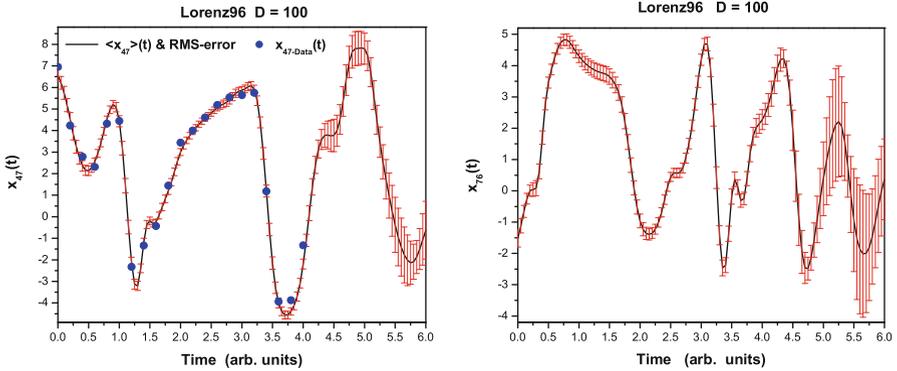


Fig. 4.8 $D = 100$ *Left* Using noisy observed data (blue circles) in the observation window $[0, 4]$, the expected value of the (observed) variable $x_{47}(t)$ (black line) was estimated using the data assimilation path integral discussed in the text. The RMS error (red error bars) around this expected value was also estimated. Using the estimated value of all 100 state variables (40 observed and 60 unobserved) along with the estimated parameter, predictions including RMS errors were made into the prediction region $[4, 6]$ as shown. Since the Lorenz96 $D = 100$ model is chaotic for the chosen forcing, the error bars grow in the prediction interval. *Right* Using observed data in the observation window $[0, 4]$, the expected value of the (unobserved) variable $x_{76}(t)$ (black line) was estimated using the data assimilation path integral discussed in the text. The RMS error (red error bars) around this expected value was also estimated. Using the estimated value of all 100 state variables (40 observed and 60 unobserved) along with the estimated parameter, predictions including RMS errors were made into the prediction region $[4, 6]$ as shown. Since the Lorenz96 $D = 100$ model is chaotic for the chosen forcing, the error bars grow in the prediction interval

4.6.2 An Example from the Lorenz96 Model with $D = 20$

For a second example we again use the Lorenz96 model, Eq. (5.28), but now with $D = 20$. We selected $f = 7.93$; again this is a value leading to chaotic orbits. We introduced noise into the dynamics and added noise to the observations presented to the model $y_l(n)$ for $L = 8$ choosing

$$l = \{0, 2, 5, 7, 9, 11, 17, 18\}. \quad (4.90)$$

The noise was taken from the Gamma distribution (Feller 1971)

$$P_{\Gamma}(x) = \frac{x^{a-1} e^{-x}}{\Gamma(a)} \quad (4.91)$$

and added to each component of the model and to the “data” generated by the model as

$$\text{scale}(P_{\Gamma}(x) - a), \quad (4.92)$$

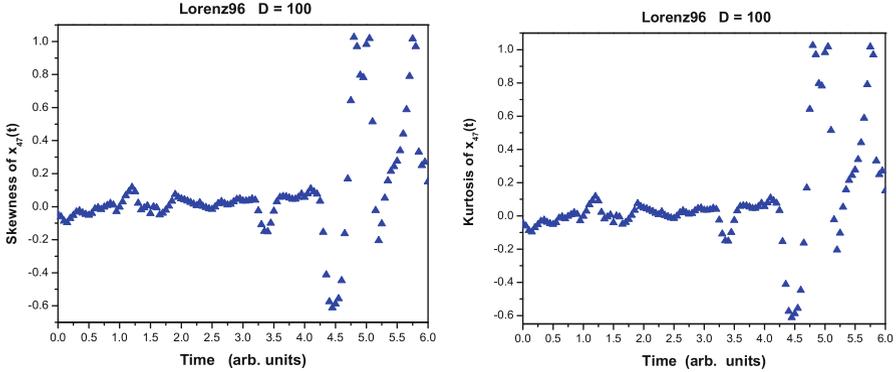


Fig. 4.9 $D = 100$ *Left* Skewness of the state variable $x_{47}(t)$, one of the observed quantities presented to the model through the data assimilation path integral. During the observation period the skewness remains small suggesting the state distribution may be approximately Gaussian in this interval. The skewness grows rapidly when observations no longer are available to guide the trajectory of the model, and it moves rapidly to its attractor which is comprised of points in 100 dimensional state space which are not distributed as a Gaussian. *Right* Kurtosis of the state variable $x_{47}(t)$, one of the observed quantities presented to the model through the data assimilation path integral. During the observation period the kurtosis remains small suggesting the state distribution may be approximately Gaussian in this interval. The kurtosis grows rapidly when observations no longer are available to guide the trajectory of the model, and it moves rapidly to its attractor which is comprised of points in 100-dimensional state space which are not distributed as a Gaussian

noting that

$$\int_0^{\infty} dx P_{\Gamma}(x) x = a. \quad (4.93)$$

We selected $a = 7$ and scale = 0.05 in the dynamical equations and scale = 0.205 in the additive noise in the data $y_l(n)$. Again we chose $R_m = 8$ and $R_f = 100$ using $m = 410$ with $\Delta t = 0.02$ in the integration to produce the data. The path integrals were of dimension 8,220, and we used 25,000 initialization Monte Carlo accept/eject steps to begin followed by 271,000 steps where statistics were recorded.

Selecting, again at random, the model error $ME_8(91)$, we display the distribution of this model error in Fig. 4.11 along with a best-fit Gaussian distribution. For this model error term, the mean value was -0.064 , while $\text{RMS}(ME_8(91))\sqrt{R_f} = 0.609$ which is quite different from the unity required for consistency with the assumption of Gaussian broadening of the deterministic $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ assumed in the path integral. One can see from Fig. 4.11 that the computed distribution $P(ME_8(91))$ is significantly narrower than a Gaussian. The same calculation but for the model error term $ME_{11}(104)$ is shown in Fig. 4.11 again along with a best-fit Gaussian. For this distribution the mean was 7×10^{-3} and $\text{RMS}(ME_{11}(104))\sqrt{R_f} = 0.64$. One can conclude that the assumption of Gaussian broadening of the transition probability used in the path integral and of Gaussian additive noise used in the conditional mutual information term of the action is not consistent with the data and the noisy model.

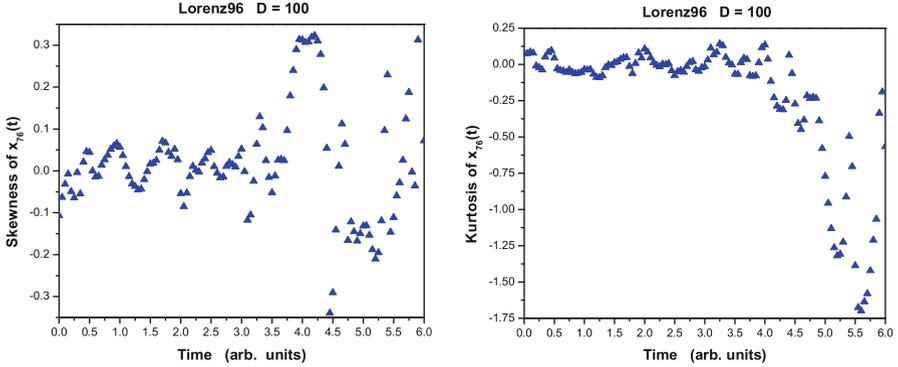


Fig. 4.10 $D = 100$ *Left* Skewness of the state variable $x_{76}(t)$, one of the unobserved quantities not presented to the model though the data assimilation path integral. During the observation period the skewness remains small suggesting the state distribution may be approximately Gaussian in this interval. The skewness grows rapidly when observations no longer are available to guide the trajectory of the model, and it moves rapidly to its attractor which is comprised of points in 100-dimensional state space which are not distributed as a Gaussian. *Right* Kurtosis of the state variable $x_{76}(t)$, one of the unobserved quantities not presented to the model though the data assimilation path integral. During the observation period the kurtosis remains small suggesting the state distribution may be approximately Gaussian in this interval. The kurtosis grows rapidly when observations no longer are available to guide the trajectory of the model, and it moves rapidly to its attractor which is comprised of points in 100-dimensional state space which are not distributed as a Gaussian

If we examine Fig. 4.12 where the expected value of the observed model variable $x_{17}(t)$ through the observation window $[0, 8.2]$ is displayed along with the calculated RMS variation about that expected value and with the observed noisy “data” points presented to the model, we see sizeable regions where the estimated value $\langle x_{17} \rangle (t)$ deviates from the observations. In Fig. 4.13 one sees very large values of the skewness and kurtosis within the observation window for the observed model variable $x_{17}(t)$ showing the departure of its distribution from a Gaussian. Figure 4.14 shows the same features for the unobserved model dynamical variable $x_{14}(t)$.

Certainly more precise statistical tests can be made to determine the deviation of the distribution of the model errors selected here from Gaussians. As Gaussians were assumed in formulating the path integral as described above, we can conclude with confidence that the Gaussian assumption is inconsistent with the data. Of course, we built this into our “twin experiment” calculation, so it is perhaps reassuring that we are able to detect this inconsistency with essentially no more effort than already required in evaluating the data assimilation path integral for our other purposes: estimation of parameters and states within and at the end of an observation period, prediction of the estimated states and their RMS errors beyond the observation window, Once we have the accepted paths $\mathbf{X}^{(j)}$, evaluation of any $F(\mathbf{X}^{(j)})$ is quite straightforward.

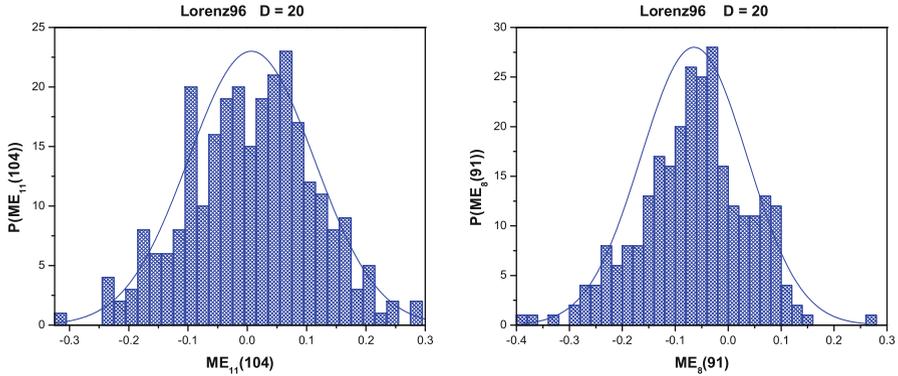


Fig. 4.11 *Left* Distribution of the stochastic model error $ME_{11}(104)$ from the $D = 20$ Lorenz96 model compared to a best-fit Gaussian distribution. *Right* Distribution of the stochastic model error $ME_8(91)$ from the $D = 20$ Lorenz96 model compared to a best-fit Gaussian distribution. In this calculation we assumed the model error was distributed as a Gaussian, when, in fact, it was drawn from a Gamma distribution. The inconsistency exhibits itself here

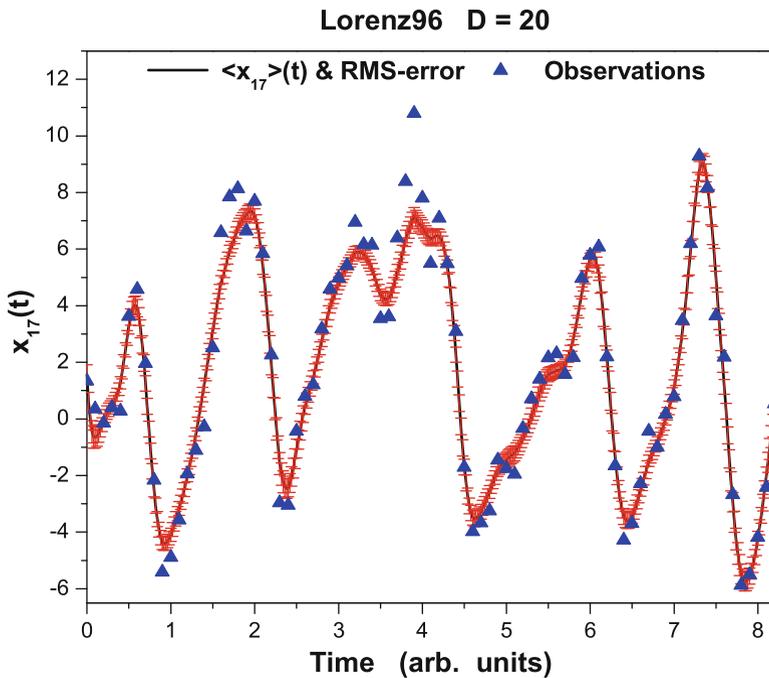


Fig. 4.12 $D = 20$ Using observed data (blue triangles) in the observation window $[0, 8.2]$, the expected value of the (observed) variable $x_{17}(t)$ (black line) was estimated using the path integral discussed in the text. The RMS error (red error bars) around this expected value was also estimated

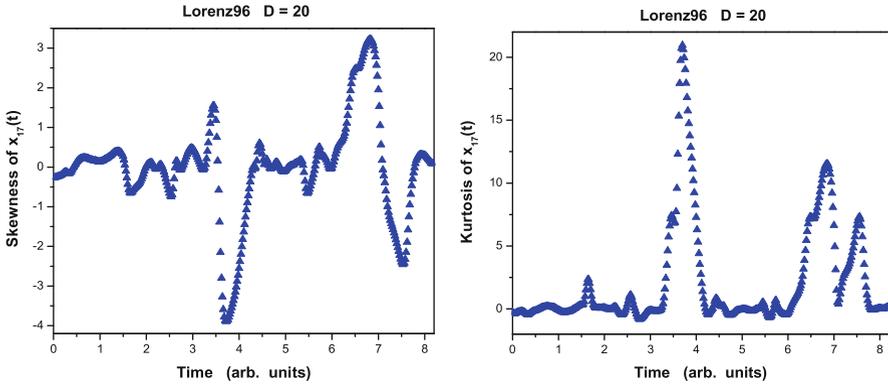


Fig. 4.13 *Left* $D = 20$ Skewness of the state variable $x_{17}(t)$, one of the observed quantities presented to the model though the data assimilation path integral. *Right* Kurtosis of the state variable $x_{17}(t)$, one of the observed quantities presented to the model though the data assimilation path integral

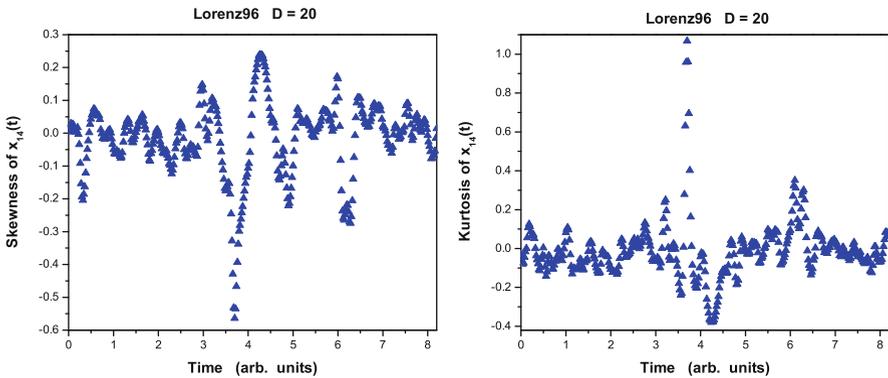


Fig. 4.14 *Left* $D = 20$ Skewness of the state variable $x_{14}(t)$, one of the unobserved quantities not presented to the model though the data assimilation path integral. *Right* Kurtosis of the state variable $x_{14}(t)$, one of the unobserved quantities not presented to the model though the data assimilation path integral

In the more interesting situation of data from field or laboratory observations, we may make the same calculations of our assumed stochastic model error terms and check equally straightforwardly for the consistency of these assumptions.

4.6.3 Comments on Consistent Model Errors

In assimilating information from measurements to a model of the observed system when the data are noisy, the models have error, and the state of the model

system uncertain when measurements begin, one must make assumptions both about the way to represent stochastic model error and noisy information transfers from the data. The latter are often known through knowledge of the sensors and the environmental noise during measurements. Model error can be structural and deterministic arising from physical processes unaccounted for in developing the model or they can be stochastic representing limits on the spatial or temporal resolution of the model or environmental noise representing fluctuations on any scale not dynamically treated in the model. The stochastic model errors broaden the manner in which the dynamics enters data assimilation. In the deterministic case the transition probability to go from $\mathbf{x}(t_n) = \mathbf{x}(n)$ to $\mathbf{x}(t_{n+1}) = \mathbf{x}(n+1)$ is $P(\mathbf{x}(n+1)|\mathbf{x}(n)) = \delta^D(\mathbf{x}(n+1) - \mathbf{f}(\mathbf{x}(n), \mathbf{p}))$. This is broadened in the stochastic data assimilation (or ensemble data assimilation) task, and an assumption on how this is represented must be made.

By comparing properties of the stochastic model errors $ME_a(n) = x_a(n+1) - f_a(\mathbf{x}(n))$ in the assumed distribution for $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ and the properties emerging from the data assimilation procedure, we tested for the consistency of the assumptions about the stochastic model errors.

Using two variants of the Lorenz96 model, we examined a case where there was demonstrable consistency of the outcome of data assimilation and assumptions about the distribution of model error, and we reported on another example where there was no such consistency.

If the distribution of model error is consistent between these two situations, this provides confidence in the precise formulation of the data assimilation tasks. Similarly, when that consistency is absent, confidence is lost. We do not provide a remedy in the case of inconsistency.

To execute this consistency test one requires basically the same numerical evaluations as in performing the overall data assimilation effort using the path integral formulation of the problem (Abarbanel 2009), so carrying out the consistency check is computationally quite inexpensive.

4.7 Synopsis and Perspectives: Evaluating the Path Integral

The data assimilation path integral permits us to address any question associated with the information contained in the distribution function of the full path \mathbf{X} along with the contribution of information from measurements passed along to the model. The conditional expected value

$$E[G(\mathbf{X})|\mathbf{Y}] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y})_{\mathbf{R}}]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y})_{\mathbf{R}}]} \quad (4.94)$$

of the function $G(\mathbf{X})$ on the path \mathbf{X} allows the formulation of any question about information in the distribution $\exp[-A_0(\mathbf{X}, \mathbf{Y})]$ in terms of a high-dimensional integral we must try to evaluate.

The various methods we have discussed for the evaluation of the integral have their individual virtues. The Laplace, saddle path method permits the use of well-established numerical optimization algorithm. It does not give corrections to the extremum path \mathbf{S}_0 where

$$\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{x}=\mathbf{s}_0} = 0, \quad (4.95)$$

and although we avoided discussing it (Zinn-Justin 2002), the expansion of the effective action, which does contain the statistical corrections to \mathbf{S}_0 , may not converge. Of course, retaining only a few terms in such a series may somehow avoid this troublesome feature.

On the positive side of an expansion about the extremum path is the possibility to determine where the saddle path approximation, or 4DVar in geophysics, fails. Always good to know, if one can.

The attractive possibility that appears from our discussion is the notion that the saddle path approximation might be bypassed altogether though the use of well-developed Monte Carlo methods reaching back to the 1950s to directly estimate the full path integral. Sometimes labeled the “Nike” method—“Just do it!”—this avoids many of the problems with convergence of saddle path approaches. It also avoids troubles in other methods (Evensen 2009) as linearizing is not required, inverting measurement functions $h_l(\mathbf{x})$ or even differentiating them is not required, and once one has identified the particular path integral that answers the question at hand, doing an integral is the issue. This actually fits well into the folklore viewpoint that in numerical estimations, estimate only that quantity of interest for a scientific reason. Do not approximate an auxiliary quantity first, for example, $\exp[-A_0(\mathbf{X})]$, then use it to calculate something else. Of course, even more important is the fact that expected values, moments about them, and marginal distributions of interesting parameters or state variables are all addressed within this context.

Nonetheless, while the “Nike” method is, from this author’s point of view, the method of choice, algorithms to perform the integral require extensive parallelism, substantial memory, and clever code development. The last word on this certainly has not been heard, though the cost in time and resources for each of these items drops annually. Another set of issues, not at all addressing in this book, are methods to accurately estimate the tails of distributions, namely, rare events. There is much to be done.

Chapter 5

Twin Experiments

Prologue to the Chapter

This chapter is devoted to the detailed discussion of several numerical simulations wherein we use a model to generate data, and then we examine how well we can use $L = 1, 2, \dots$ of the time series for state variables of the model to estimate fixed parameters within the model and the time series of the state variables not presented to or known to the model. These are “twin experiments” and have often been used to exercise the methods one adopts for approximating the path integral for the statistical data assimilation problem.

We will look at three examples: (1) Hodgkin–Huxley neuron models, (2) the Lorenz96 geophysical model, and (3) the more complex driven shallow water equations on a rotating plane. We will find that the twin experiments serve not only the “testing of methods” purpose, but they allow one to consider design of experiments which may be utilized in testing the model one proposes for describing the object of our observations.

From these considerations will come a strategy for developing a large, many degrees of freedom, many intrinsic current neuron model that will be broadly useful in analyzing electrophysiological data from many classes of neuron. The data used via our data assimilation approach will prune away parts of the model not required by the observations.

Further, and perhaps more broadly of significance, we show how one may use the model, absent any observed data, in a twin experiment mode to establish how many measurements at each observation time must be acquired to allow the estimation procedures in approximating the path integral to work well. In evaluating the quality of the model, prediction beyond the observation window used to complete the model is the key metric. Using the method of “data assimilation without data” explored in Chap. 3 then gives us the tool for that metric.

5.1 The Roles of Twin Experiments

We have often used the idea of twin experiments in our considerations. At first glance twin experiments, in which we generate the “data” using a known model, are just ways of testing methods for statistical data assimilation. In a twin experiment, we know everything about the model, and we know everything about the data. We may wish to investigate how well our methods for estimation of the underlying path integral work on the raw data produced by the model or on model generated data contaminated by noise with some distribution or even on models similar to our original selection but with some model errors introduced. We also know the data for both the observed and unobserved state variables allowing comparisons both of the quality of estimations for parameters and for estimates of the unobserved states. Clearly the latter is unavailable in laboratory or field experiments.

Twin experiments are also very important as they allow us to address questions such as how many measurements are required for the accuracy of state and parameter estimations. They can be used to identify which measurements to make. They permit us to ask how frequently in time this number of measurements should be performed in order to make accurate estimations of model parameters and model states. Twin experiments are also useful as a design tool for experiments as they may indicate properties of the stimulus or forcing of the experimental system in order to explore the full dynamical range of its response.

We will now discuss a series of twin experiments in several areas of interest. In particular, we look at biophysical questions involving individual neurons and networks of neurons. Then we have a look at geophysical problems where the questions are similar and the analysis is also similar. One purpose of asking essentially the same questions in two apparently vastly different arenas of scientific inquiry is to illustrate the broad applicability of the methods and the questions.

Twin experiments can be used then as a design tool in selecting experiments which probe the full dynamical range of the models we expect to use for those experiments. In both lab and field observations when constraints on the instruments or immutability of the available forcing functions restrict the stimuli one can utilize, one can anticipate where the knowledge of parameters and states might be inadequate or weak. In our exploration of twin experiments and subsequently in the analysis of laboratory experiments in a neurobiological context we will see that the design of the experiments is strongly assisted by twin experiments in simulation.

5.2 Neuron Models

Our first investigation of some twin experiments comes in the context of making, testing, and validating biophysically based models, within the Hodgkin–Huxley–Katz formulation of the dynamics of these systems (Johnston and Wu 1995). This formulation was developed in the 1930–1950s by a group of investigators in Cambridge in the UK. The core idea was to describe in a phenomenological manner

the way ion currents are conducted by voltage dependent permeabilities across the bi-lipid membranes of cells. The actual physical channels through which the ions pass are complex proteins, of significant interest in themselves. These proteins penetrate the cell membrane with physical extensions both into the intracellular and the extracellular medium.

The key idea developed by the investigators some decades ago is that there is a balance between the different concentration of ions within and without the neuron cell associated with the diffusion of ions to lower concentrations and the electrostatic forcing of the same ions associated with the difference in potential across the cell membrane—a capacitor able to separate charges. This balance defines the Nernst potential, known long before the Cambridge investigators, and is a consequence of equilibrium statistical mechanics (Johnston and Wu 1995).

The innovative ingredient was to recognize that the voltage-dependent permeability of the protein-based ion channels allowed the quantitative description of an ion current as

$$I_{\text{ion}}(t) = g_{\text{ion}} \text{activation}(t)^{n1} \text{inactivation}(t)^{n2} (E_{\text{Nernst}} - V_{\text{cell}}(t)), \quad (5.1)$$

where g_{ion} is a conductance identified as the maximum allowed when the ion channels are open. $V_{\text{cell}}(t)$ is the potential across the cell membrane. $n1$ and $n2$ are integers.

The dimensionless time-dependent kinetic variables $\text{activation}(t)$ and $\text{inactivation}(t)$ are phenomenological state variables describing the opening and closing of the protein channel associated with the ion in question; they lie in the range $[0, 1]$. These are taken to satisfy linear kinetic equations, as we will describe soon. The integer powers $p1, p2$ are to be determined through experiment. The “reversal potential” E_{Nernst} is where the diffusion and electrostatic forces on the ions are balanced and from nineteenth century kinetic theory considerations is determined by the intracellular ion concentration $[\text{Ion}]_i$ and the extracellular ion concentration $[\text{Ion}]_e$ as

$$E_{\text{Nernst}} = \frac{RT}{zF} \log \left[\frac{[\text{Ion}]_e}{[\text{Ion}]_i} \right], \quad (5.2)$$

where $R = 8.314 \text{ J/deg-K mol}$ is the gas constant, T is the temperature in degrees Kelvin, z is the charge of the ion in units of the electron charge, and $F = 96485 \text{ C/mol}$ is the Faraday constant. For $z = 1$ and $T = 37^\circ\text{C}$, $\frac{RT}{zF} = 26.15 \text{ mV}$, setting the scale of neuron electrical activity.

5.2.1 NaKL Hodgkin–Huxley Model

We return to the NaKL HH neuron model discussed in Chaps. 2 and 3 to illustrate some of the ways in which twin experiments can be utilized in experimental design as well as guiding statistical data assimilation in the analysis of experiments.

The Hodgkin–Huxley (HH) dynamical equations for the NaKL neuron consist first of a current conservation statement for the cross membrane voltage $V(t)$:

$$C \frac{dV(t)}{dt} = g_{\text{Na}} m(t)^3 h(t) (E_{\text{Na}} - V(t)) + g_{\text{K}} n(t)^4 (E_{\text{K}} - V(t)) + g_{\text{L}} (E_{\text{L}} - V(t)) + I_{\text{app}}(t), \quad (5.3)$$

where the g_{Na} , g_{K} , and g_{L} are maximal conductances and the E_{Na} , E_{K} , and E_{L} are reversal potentials for the Na, K, and leak channels. $I_{\text{app}}(t)$ is an applied external current selected in an experiment.

The gating variables $a_i(t) = \{m(t), h(t), n(t)\}$ are discussed in many textbooks and reviews (Johnston and Wu 1995; ModelDB 2012; Graham 2002). Each satisfies a first-order kinetic equation of the form

$$\frac{da_i(t)}{dt} = \frac{a_{i0}(V(t)) - a_i(t)}{\tau_i(V(t))}. \quad (5.4)$$

The kinetic terms $a_{i0}(V)$ and $\tau_i(V)$ are taken here in the functional form

$$\begin{aligned} a_{i0}(V) &= \frac{1}{2} \left(1 + \tanh \left[\frac{(V - v_a)}{dva} \right] \right) \\ \tau_i(V) &= t_{a0} + t_{a1} \left(1 - \tanh^2 \left[\frac{(V - vat)}{dvat} \right] \right) \quad \text{or} \\ \tau_i(V) &= t_{a0} + t_{a1} \tanh \left[\frac{(V - vat)}{dvat} \right]. \end{aligned} \quad (5.5)$$

The NaKL model has four state variables and 25 fixed parameters.

The applied current waveform was taken from the output of a chaotic system which has a Fourier spectral content providing frequencies in a band approximating 50% of the frequencies in the neuron model voltage output. The time scale of the current relative to the time step for the integration of the model is called ρ_I . When ρ_I is large, the frequency of variations in the stimulating current is large compared to the voltage response of the neuron. In this case, much of the information in the forcing signal $I_{\text{app}}(t)$ is filtered out by the RC time constant of the cell membrane acting as a capacitor. These dynamical equations are used, in a twin experiment, to produce data $\{V_{\text{data}}(t), m_{\text{data}}(t), h_{\text{data}}(t), n_{\text{data}}(t)\}$. Only $V_{\text{data}}(t)$ is presented to the model in our twin experiments.

In Sect. 2.3 we examined the accuracy with which one can estimate the model parameters and unobserved states using a variational principle with no model errors. There we also looked at the ability of the variational methods to accurately estimate the full state of the neuron model at the end of observations by using the model to predict. Here we wish to look at other aspects of the data assimilation procedure as illustrated by twin experiments.

5.2.2 The Importance of the Regularizing Variable $u(t)$

The variational principle for the NaKL neuron with no model error uses data $y(t) = V_{\text{data}}(t)$ for the membrane voltage generated by the model itself. This “observed” voltage was presented to the model dynamical equations by being introduced into the voltage equation (5.3) using a coupling term $u(t)(y(t) - V(t))$:

$$\frac{dV(t)}{dt} = \frac{1}{C} \left\{ g_{\text{Na}}m(t)^3h(t)(E_{\text{Na}} - V(t)) + g_{\text{K}}n(t)^4(E_{\text{K}} - V(t)) + g_{\text{L}}(E_{\text{L}} - V(t)) + I_{\text{app}}(t) \right\} + u(t)(y(t) - V(t)). \quad (5.6)$$

At the same time a cost for the use of $u(t)$ was introduced into the cost function for the deviation of the observations and the model output as

$$\sum_{n=0}^m (y(t_n) - V(t_n))^2 \rightarrow \sum_{n=0}^m \{(y(t_n) - V(t_n))^2 + u(t_n)^2\}. \quad (5.7)$$

The importance of the use of the control term $u(t)$ in the numerical accuracy of the state and parameter estimations was assessed by performing the calculation that led to Table 2.4 with precisely the same model, the same data, the same stimulus, and the same algorithm for nonlinear estimation, but setting $u(t) = 0.0$ throughout. This result is presented in Table 5.1 where we see clearly that the overall quality of the estimation is degraded. This suggests that the use of the control variable $u(t)$ during the data assimilation will prove a good strategy even when the impediments of chaotic behavior are absent. The role of the added “control” term $u(t)(y(t) - V(t))$ is to direct the solution to the differential equation for $V(t)$ to the values observed in the data $y(t)$. This is seen to be helpful even when it is not necessary when chaos is absent.

This is a good time to remind ourselves that $u(t)$ has no physical meaning. It is used solely as a regularizing variable to address instabilities on the synchronization manifold $y_{\text{data}}(t) \approx h(\mathbf{x}(t))$ and as a numerical helper to direct the model output on the observed states to the information provided by the observed data.

Since $u(t)$ is absent at the end of the calculation, as evidenced by the values of $R(t)$ in Fig. 2.17, its utility in guiding the numerical optimization seems clear. There is a computational cost as the values of $u(t_n)$ are part of the search procedure. In addition, though not shown here, one can track the values of $u(t_n)$ as the numerical optimization routine proceeds. In the initial stages of the optimization task, $u(t_n)$ can become quite large, yet as the data synchronizes with the model output, $V(t) \rightarrow y(t)$, one sees that the term $u(t_n)^2$ in the cost function efficiently drives the $u(t_n)$ to zero within machine roundoff error.

When $u(t)$ is absent altogether, there is no direct communication between the data $y(t)$ and the corresponding model output $V(t)$ in the dynamical equations. The

Table 5.1 NaKL model neuron parameters in data and estimated parameters in NaKL HH model; $\rho_I = 0.5$

Parameter name	Parameter value in the data	Estimated value of the parameter
g_{Na}	120 mS/cm ²	100.0 mS/cm ²
E_{Na}	50.0 mV	50.71 mV
g_K	20 mS/cm ²	10.0 mS/cm ²
E_K	-77.0 mV	-115.0 mV
g_L	0.3 mS/cm ²	0.262 mS/cm ²
E_L	-54.4 mV	-75 mV
$v_m = v_{mt}$	-40.0 mV	-57.9 mV
$dv_m = dv_{mt}$	15.0 mV	0.378 mV
t_{m0}	0.1 ms	0.2368 ms
t_{m1}	0.4 ms	7.34e-14 ms
$v_h = v_{ht}$	-60.0 mV	-58.13 mV
$dv_h = dv_{ht}$	-15.0 mV	-0.47 mV
t_{h0}	1.0 ms	0.48 ms
t_{h1}	7.0 ms	3.43 ms
$v_n = v_{nt}$	-55.0 mV	-63.06 mV
$dv_n = dv_{nt}$	30.0 mV	55.0 mV
t_{n0}	1.0 ms	1.00 ms
t_{n1}	5.0 ms	0.285e-12 ms

No coupling of observed data into the dynamical equations was used, so, $u(t) = 0.0$. In all other aspects this is the same calculation as displayed in Table 2.4

search through the state variables $\mathbf{x}(t) = \{V(t), m(t), h(t), n(t)\}$ and parameter space might indeed proceed to find excellent estimates for these state variables in the observation window and for the parameters; however, the same calculation done with $u(t)$ present and without $u(t)$ present results in much worse estimates in the latter case as seen in Table 5.1. It seems clear that information about the data is of significant value when it goes beyond that contained only in the cost function. We have also argued that when the synchronization manifold is unstable, as in the observation of chaotic data, the role of the regularizing variable $u(t)$ is actually essential for the optimization when no model errors are assumed.

There are other regularization approaches. One idea introduces the data $y(t)$ into the dynamical equations of the model as $V(t) \rightarrow U(t)y(t) + (1 - U(t))V(t)$ with a similar addition to the cost function. In this format $U(t)$ is dimensionless and $0 \leq U(t) \leq 1$. When $U(t) = 0$, no regularization occurs. In the opposite case, if we set $U(t) = 1$, however, we would have eliminated the dynamical variable $V(t)$ entirely from the model and reduced the dynamical equations to $D - 1 = 3$, here. This is the strategy for synchronization introduced by Pecora and Carroll (1990). It eliminates the observable from the model equations altogether and injects measurement information into the remaining equations for the unobserved variables.

Table 5.2 Parameters in “data” and estimated parameters in NaKL HH model; $\rho_I = 10.0$

Parameter name	Parameter value in the data	Estimated value of the parameter
g_{Na}	120.0 mS/cm ²	100.0 mS/cm ²
E_{Na}	50.0 mV	50.0 mV
g_K	20.0 mS/cm ²	14.6 mS/cm ²
E_K	-77.0 mV	-115.0 mV
g_L	0.3 mS/cm ²	1.23e-13 mS/cm ²
E_L	-54.4 mV	-58.22 mV
vm = vmt	-40.0 mV	-47.78 mV
dvm=dvmt	15.0 mV	30.0 mV
t_{m0}	1.0 ms	0.636 ms
t_{m1}	0.4 ms	3.1 ms
vh=vht	-60.0 mV	-83.72 mV
dvh=dvht	-15.0 mV	-6.5 mV
t_{h0}	1.0 ms	1.602 ms
t_{h1}	7.0 ms	12.0 ms
vn=vnt	-55.0 mV	-52.04 mV
dvn=dvnt	30.0 mV	13.80 mV
t_{n0}	1.0	7.25 ms
t_{n1}	5.0 ms	15.0 ms

In all other aspects this is the same calculation whose results are reported in Table 2.4

5.2.3 Frequency Content of the Stimulus $I_{app}(t)$

Proper estimation of the parameters of the model will depend on providing a stimulus current that adequately explores regions of the state space where different subsets of the channels are opened or closed. If the frequency content of the stimulus is such that it fails to spend sufficient time in these dynamical regions of model variability, the assimilation results will be degraded. We examined this by increasing the relative frequency of the stimulus current ρ_I by selecting the same waveform for $I_{app}(t)$ and sampling it more frequently. In the estimations examined earlier ρ_I was 0.5; namely, it was relatively slow compared to the dynamics of the model. Good estimates hold for ρ_I up to about 5.0. At $\rho_I \approx 10.0$, these estimates became of low quality, Table 5.2.

In a twin experiment we are also able to examine estimates of the voltage as well as estimates of the gating variables over the observation window. In Fig. 5.1 **Left**, the estimated membrane voltage is compared to the known $V_{data}(t)$ when $\rho_I = 10.0$. This alone is misleading; however, as seen when one of the unobserved state variables, $m(t)$ is compared to its estimated values as shown in Fig. 5.1 **Right**. By changing ρ_I from 0.5, as used in Chap. 2, to 10.0, as used in this example, we have lost the capability of the model to transfer information from the observations to an unobserved state variable.

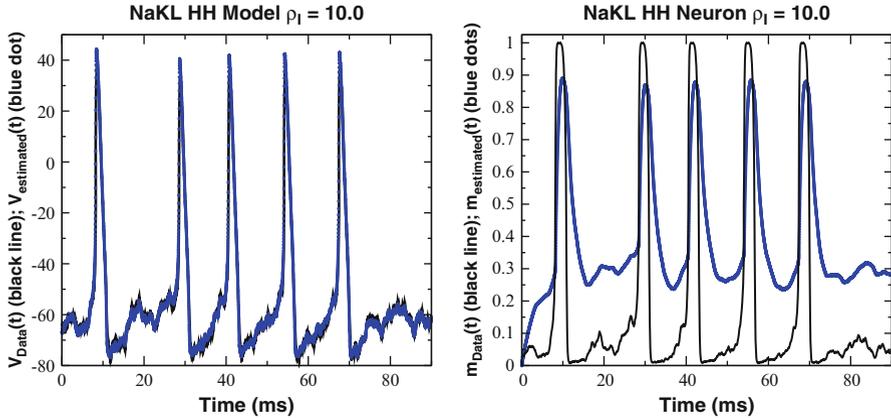


Fig. 5.1 NaKL model. *Left* We display the membrane voltage $V_{\text{data}}(t)$ from “data” generated in a twin experiment using the NaKL model as a data source. Also we display the estimated membrane voltage $V_{\text{estimated}}(t)$ using the saddle path variational method with no model errors. The frequency scale of the stimulus relative to the neuron response is $\rho_I = 10.0$. The apparent good estimation of the voltage results from the large magnitudes of the control $u(t)$. *Right* We display the Na^+ activation gating variable $m_{\text{data}}(t)$ from “data” generated in a twin experiment using the NaKL model as a data source. Also we display the estimated membrane voltage Na^+ activation gating variable $m_{\text{estimated}}(t)$. The mismatch between the estimated $m(t)$ and the known $m(t)$ demonstrates one of the uses of a twin experiment, here suggesting we have used a stimulus with too much high frequency content. The frequency scale of the stimulus to the neuron response is $\rho_I = 10.0$. In a laboratory experiment, $m(t)$ would be an unobserved state variable

5.2.4 Additive Noise in the Observations

Until now we have allowed only numerical roundoff error to contaminate the “observations” $y(t) = V_{\text{data}}(t)$. To assess the role of noise in measurements, which will eventually degrade any state and parameter estimation procedure, we added noise drawn from a uniform distribution to the data $y(t)$ generated from our NaKL model. The noise level selected reduces the signal-to-noise ratio to 30 dB. As shown in Fig. 5.2, the added noise does not compromise the data assimilation procedure. In fact, the estimated voltage is much closer to the original “clean” data than to the noisy data presented to the model. We observe that for modestly noisy data our approach succeeds in resolving a clean signal contaminated by instrumental or environmental addition of noise. Calculations suggested that until the signal-to-noise ratio decreased to about 20 dB, the procedure remained successful.

In neurobiological experiments making intracellular recordings, a noise floor in the range of 0.5 mV is commonly achieved. Spiking activity results in signals of up to 100 mV and certainly at least 40–60 mV so the signal-to-noise ratio required for accurate modeling is easily achieved. Smaller potentials representing input to a cell from other cells are often in the range of a few mV, however, so this may represent a signal-to-noise ratio problem when we extend these ideas to modeling interactions among neurons in a network.

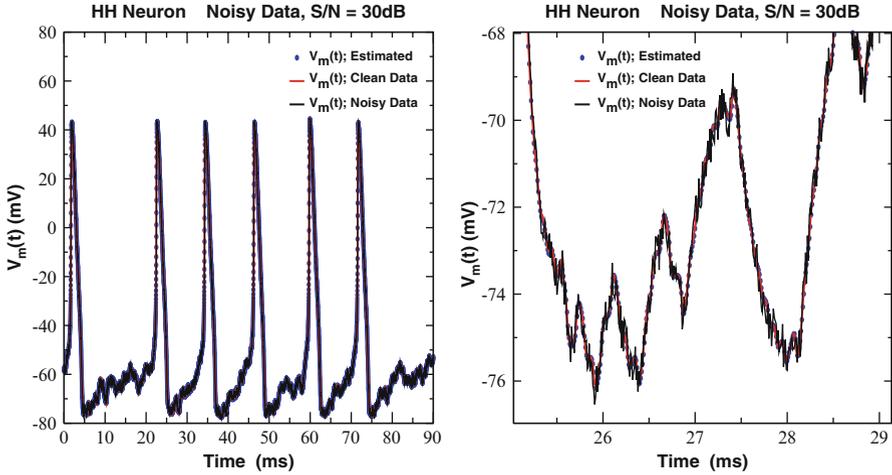


Fig. 5.2 *Left panel* Estimated membrane voltage (blue dots) along with original clean data (red line) and the noisy voltage data (black line) presented to the NaKL model in the dynamical data assimilation approach. Here the signal-to-noise ratio is 30 dB, and, as above, $\rho_I = 0.5$. *Right panel* A blowup of a section of the left panel showing how the estimation procedure has removed much of the noise added to the original clean data. The model acts here as a nonlinear noise filter. See Abarbanel et al. (2011) for a longer discussion of this effect

When noise levels grow too large, any estimation protocol will fail. To examine how robustly the analysis of observed and unobserved state variables acts as noise levels are increased, we evaluated the estimation error in each of the state variables as the noise level in the voltage data was increased reducing the signal-to-noise ratio from 50 to 10 dB. The correlation of the known data $\{y_1(t), y_2(t), y_3(t), y_4(t)\}$ with the model output when additive noise of a specified level was presented to the model is shown in Fig. 5.3. We evaluated the mean value of the data $\langle y_a \rangle$ and the model output $\langle x_a \rangle$

$$\begin{aligned} \langle y_{a\text{-clean}} \rangle &= \frac{1}{N} \sum_{n=0}^{N-1} y_{a\text{-clean}}(t_n) \\ &\text{and} \\ \langle x_{a\text{-clean}} \rangle &= \frac{1}{N} \sum_{n=0}^{N-1} x_{a\text{-clean}}(t_n), \end{aligned} \tag{5.8}$$

when no noise was added to the data $y_{a\text{-clean}}(t_n)$ and the model output was $x_{a\text{-clean}}(t_n)$. We then evaluated the RMS difference between $y_a(t_n) - \langle y_{a\text{-clean}} \rangle$ and $x_a(t_n) - \langle x_{a\text{-clean}} \rangle$ normalized by $\langle x_{a\text{-clean}} \rangle$ for $a = 1, \dots, 4$ as a function of the signal-to-noise ratio in the data. The accuracy of the estimates for low noise, signal-to-noise (S/N) levels of about 30 dB or more, is seen to be quite high. As the S/N ratio falls below this, the quality of the estimates remains acceptable until one reaches about S/N less than 10 dB below which (not shown) the estimation error rises rapidly.

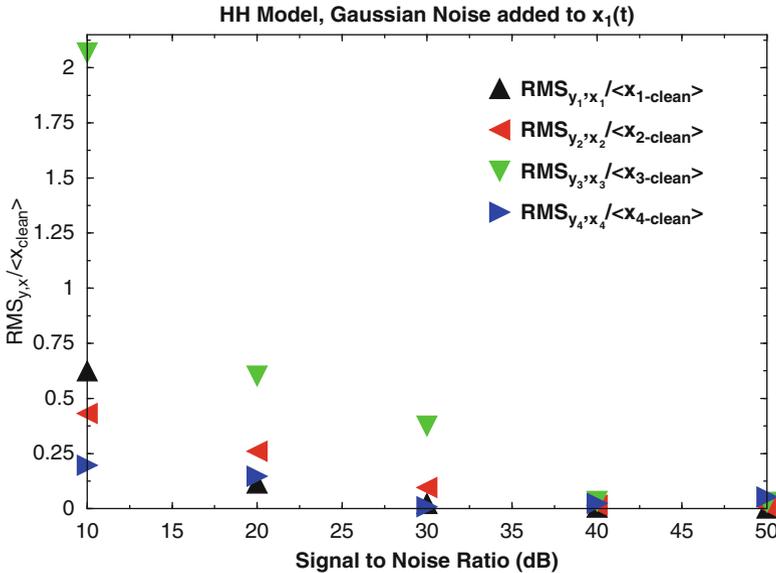


Fig. 5.3 Estimation error for the NaKL HH model when noise of various levels is added to the data. The estimation procedure is rather robust to increasing noise levels until the signal-to-noise level falls below about 10 dB (not shown). $\text{RMS}_{x_a, y_a} =$

$$\sqrt{\frac{1}{m+1} \sum_{n=0}^m \left[(y_a(n) - \langle y_{a\text{-clean}} \rangle) - (x_a(n) - \langle x_{a\text{-clean}} \rangle) \right]^2}$$

5.2.5 An Additional Current: I_h

Neurons express many voltage-gated channels in addition to the Na and K channels directly responsible for action potential generation (Graham 2002; Hille 1992; ModelDB 2012). Determining which channels are expressed in a population of neurons typically requires molecular or pharmacological data (Johnston and Wu 1995), but the data assimilation procedure we describe here has the potential to infer which channels contribute to a neuron’s response properties without the pharmacological procedures.

To explore this we expanded the NaKL model to include another current, the I_h current (Huguenard and McCormick 1992), which has slower kinetics than the other channels in the model and is activated by voltages near the rest voltage of the neuron, namely, by hyperpolarization. We call this the NaKLh neuron model.

Including an additional current in the model allows us to address the interesting questions:

- If data from a more complex model including I_h (NaKLh) is presented to a simpler NaKL model, are the parameter estimates in the simpler model still accurate?

Table 5.3 Parameters for I_h used in the NaKLh Hodgkin–Huxley Model

Name	Value	Name	Value
g_h	1.21 mS/cm ²	t_{hc0}	0.1 ms
E_h	−40.0 mV	t_{hc1}	193.5 ms
vhc	−75.0 mV	vhct	−80 mV
dvhc	−11.0 mV	dvhct	21.0 mV

This neuron has these currents: Na, K, Leak, and I_h . The parameters for the Na, K, and Leak channels have been given earlier

- If data from a model that lacks an I_h or other current (NaKL) are presented to a more complex model (NaKLh), will the dynamical data assimilation procedure inform us that I_h is absent?

In the first instance as the I_h current and its degrees of freedom are absent in the NaKL model of the data, one is projecting the state space of a more complex system into the smaller space of the NaKL model. One is asking the smaller number of states in NaKL to exhibit the dynamics of the larger data source. In this circumstance we expect some signals of the inability of the model to represent the data.

In the second instance we know that the data from the NaKL model does not include the state activity of the I_h current, so we would expect that the NaKLh model would “shed” its I_h dynamics and respond only in the subspace of the overall model space utilizing the NaKL degrees of freedom. The I_h current was represented by an additional term in the Hodgkin–Huxley voltage equation,

$$I_h(t) = g_h h c(t)(E_h - V(t)), \quad (5.9)$$

as well as an additional equation for the dynamics of the I_h gating variable:

$$\frac{dhc(t)}{dt} = \frac{hc_0(V(t)) - hc(t)}{\tau_{hc}(V)} \quad (5.10)$$

$$hc_0(V) = \frac{1}{2} \left(1 + \tanh \left[\frac{(V - \text{vhc})}{\text{dvhc}} \right] \right) \text{ and } \tau_{hc}(V) = t_{hc0} + t_{hc1} \tanh \left[\frac{(V - \text{vhct})}{\text{dvhct}} \right].$$

5.2.6 NaKLh Twin Experiments

Using the NaKLh model and the parameter values for Na, K, and Leak channels given before and those in Table 5.3 for I_h , data for $\{V(t), m(t), h(t), n(t), hc(t)\}$ were generated. We now have five state variables and 26 unknown parameters in the model, and we present only $V_{\text{data}}(t)$ to the NaKLh model to achieve this goal.

We used two different stimulation currents, with the same ρ_I and waveform shape as earlier, but with different amplitudes. The “strong” stimulus was 1.85 times the amplitude of the “weak” stimulus. The states and parameters of the NaKLh model were estimated using the variational principle with no model errors separately for the data from each of these stimuli. As shown in Table 2.4 the estimates for all the

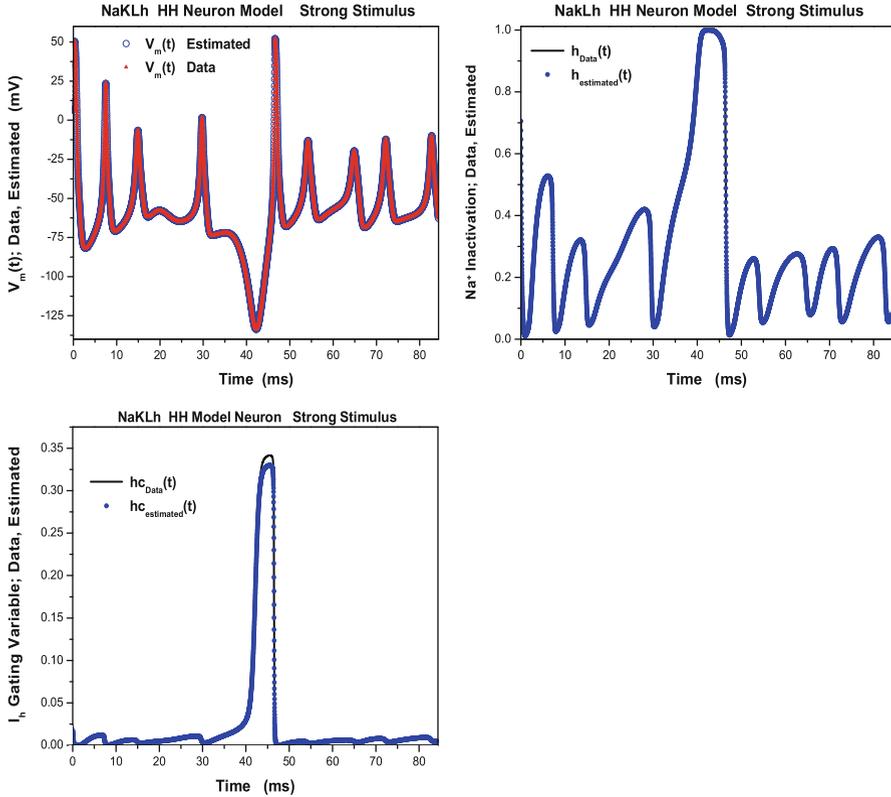


Fig. 5.4 NaKLh twin experiments: strong stimulus. *Top left* Estimated membrane voltage (blue circles) along with original data (red triangles) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “strong” stimulus which excites both depolarized state variables and hyperpolarized state variables. *Top right* Estimated Na⁺ inactivation gating variable (blue circles) along with original data (black line) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “strong” stimulus which excites both depolarized state variables and hyperpolarized state variables, as above. *Bottom* Estimated I_h activation gating variable (blue circles) along with original data (black line) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “strong” stimulus which excites both depolarized state variables and hyperpolarized state variables $\rho_I = 0.5$

parameters in the NaKL model are accurate using either stimulus, but only the data for the strong stimulus gives good estimates of the parameters for the I_h current. This suggests that the weaker stimulus did not adequately explore the dynamics associated with the I_h current. As shown in Fig. 5.4, the estimated voltage and state variables are good estimates for all the currents when the stimulus is strong. When the stimulus is weak, the estimated voltage and state variables for the NaKL model are still good, but the estimated state associated with the I_h current is represented poorly (Fig. 5.4).

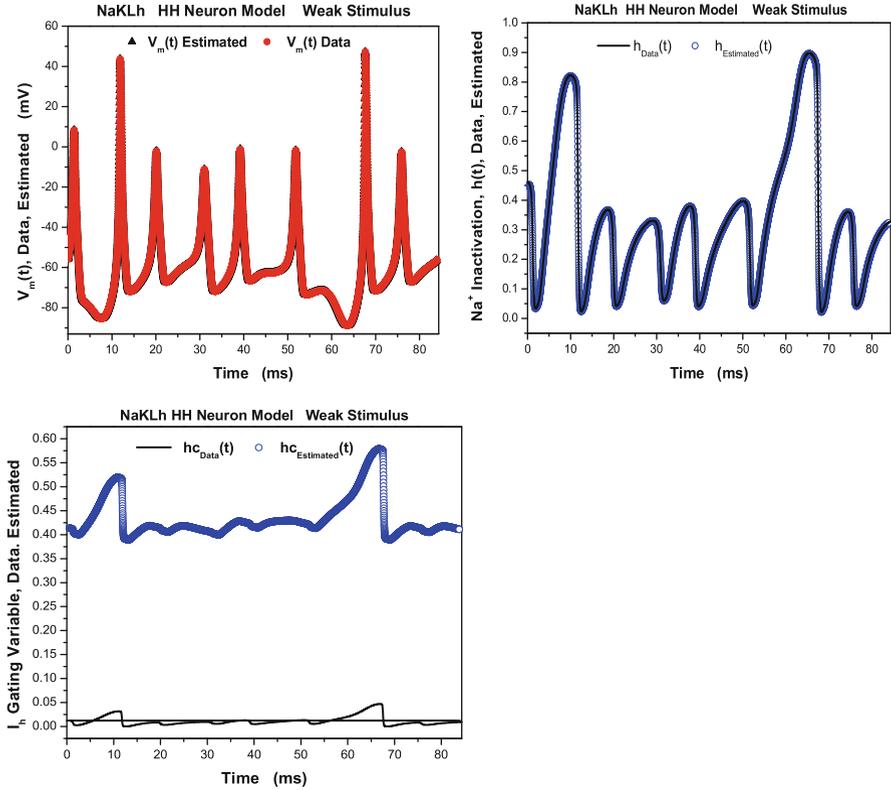


Fig. 5.5 NaKLh twin experiments: weak stimulus. *Top left* Estimated membrane voltage (black triangles) along with original data (red circles) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “weak” stimulus which excites polarized state variables. The “weak” stimulus provides sufficient excitation to the Na and K channels to allow accurate estimation of their properties. *Top right* Estimated Na⁺ inactivation gating variable (blue circles) along with original data (black line) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “weak” stimulus which excites polarized state variables. The “weak” stimulus provides sufficient excitation to the Na and K channels to allow accurate estimation of their properties. *Bottom* Estimated I_h activation gating variable (blue circles) along with original data (black line) as presented to the NaKLh model. The stimulus current was taken as usual from a chaotic system. This result uses the “weak” stimulus which excites depolarized state variables. The h-current is not well stimulated by this “weak” $I_{app}(t)$. $\rho_I = 0.5$

In an experimental setting, however, only the quality of the estimation of the voltage trace (Fig. 5.5) and the $R(t)$ consistency test would be available. Neither of these indicate that the weaker stimulus fails to excite one of the currents in the model, though the latter may well indicate an inconsistency of the model with the data. Thus, an additional test for model validity is needed, and we look to the ability of the model to predict beyond the window where observations are made.

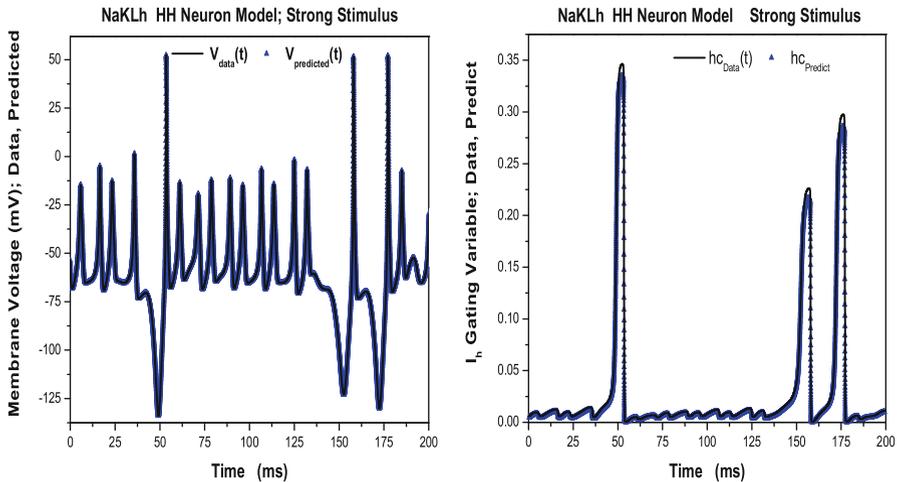


Fig. 5.6 *Left* Prediction of the membrane voltage for times greater than the observation window with state variables and parameters estimated with the strong stimulus. *Right* Prediction of the I_h gating variable $hc(t)$ for times greater than the observation window with state variables and parameters estimated with the strong stimulus. In both cases, the values of the current's parameters and state variables are accurately estimated

5.2.7 Model Testing Through Prediction

An important piece of a model investigation we have not yet covered in our discussion is the ability of the model neuron with parameters and state variables estimated in an observation window to predict the response of that neuron to new stimuli. If this is successful for a class of stimuli that one expects when the neuron is placed into a network, we can anticipate that it will respond properly in the network. We examined this issue in the context of our two stimuli: weak and strong, knowing the former was not giving sufficient stimulation of the h -current.

To validate the models' consistency when presented with weak and strong stimuli, we tested their ability to predict responses to the other stimulus. Using the parameters estimated with data in the observation window, we employed the estimated state variables $\{V(T), m(T), h(T), n(T), hc(T)\}$ to predict the behavior of the model neuron for $t > T$. When the model resulting from estimation on the strong stimulus was tested on the weak stimulus, we found that the model was able to predict the response of the neuron to the weak stimulus with a high degree of accuracy (Fig. 5.6, **Left**). In this twin experiment we are also able to compare the predicted values for the state variables against their true values (Fig. 5.6, **Right**), which display a similar degree of accuracy.

Completing the NaKlh model using the weak stimulus left us unable to accurately predict the response to the strong stimulus. Both the phase and amplitude of the predictions are degraded, especially at times when the stimulus was strongly

hyperpolarizing. Thus, because the weak stimulus did not give sufficient information about I_h , the predictions are inadequate. Predictive experiments in a laboratory setting are a useful way to test whether a candidate stimulus in conjunction with a neuron model is capable of adequately probing the model in question. These results also emphasize the importance of using model predictions to cross-validate them against different kinds of stimuli.

Perhaps a word about the word validation is in order. There is a mathematical sense in which it means prove. Perhaps it did just that to the Romans, but possibly we can recognize another understanding. We are employing the notion of validation as that of confirming or qualifying. We require a metric for the quality of a model, and there is surely no single way we can achieve that. We adopt the idea that prediction is the metric to stand for practical validity of a model. Given all models are wrong, presumably we could never validate any model. It then makes some sense to let the notion of validity represent confirmation or qualification of a model embodying the dynamical action of the source of data.

5.2.8 *Robustness to Model Errors; NaKL Model \leftrightarrow NaKLh Model*

We want to contrast what happens when we present data from the NaKLh model first to itself, a now-familiar twin experiment, to what happens when we present NaKLh voltage data to the NaKL model. The latter presents the data to a wrong model, namely, a model missing the I_h degrees of freedom.

NaKLh Model \rightarrow NaKLh Model

First we take the NaKLh model with a strong stimulus and generate “data” without added noise. We present this NaKLh model output voltage along with the same injected current stimulus to an NaKLh model. Using our variational principle with no model errors we are able to accurately estimate the parameters, as reported in Table 5.4. As a further confirmation of the method in Fig. 5.7, **Left**, we also display the voltage response in this circumstance. We confirm this consistency of the model with the data by evaluating $R(t)$ as shown in Fig. 5.8. Recall that $R(t) \approx 1$ means the dynamical model is consistent with the data presented to it. We further examine the accuracy of our estimation of an unobserved state variable $m(t)$ for the NaKLh model. This is shown in Fig. 5.7, **Right**. As in the NaKL \rightarrow NaKL twin experiment, we see that our assimilation methods using the variational approximation to the path integral in the case of no model error works rather well.

Table 5.4 Parameters in an NaKLh HH model where the maximal conductance of the I_h current from two different amplitudes of the stimulus current

Name of parameter	“Data” value	Estimated value strong stimulus	Estimated value weak stimulus
g_{Na}	120 mS/cm ²	120.273 mS/cm ²	119.884 mS/cm ²
E_{Na}	55.0 mV	54.980 mV	54.9933 mV
g_K	20 mS/cm ²	20.1476 mS/cm ²	20.1480 mS/cm ²
E_K	-77.0 mV	-77.01318 mV	-77.0192 mV
g_L	0.3 mS/cm ²	0.306754 mS/cm ²	0.298518 mS/cm ²
E_L	-54.4 mV	-54.7061 mV	-54.1206 mV
vm = vmt	-34.0 mV	33.98345 mV	-33.9948 mV
dvm=dvmt	34.0 mV	33.9958 mV	33.9847 mV
t_{m0}	0.01 ms	0.010299 ms	0.010133 ms
t_{m1}	0.5 ms	0.500221 ms	0.500495 ms
vh=vht	-60.0 mV	-59.99098 mV	-59.9611 mV
dvh=dvht	-19.0 mV	-18.98205 mV	-18.9729 mV
t_{h0}	0.2 ms	0.198847 ms	0.199929 ms
t_{h1}	8.5 ms	8.51632 ms	8.50784 ms
vn=vnt	-65.0 mV	-64.83444 mV	-64.9436 mV
dvn=dvnt	45.0 mV	44.9664 mV	45.1465 mV
t_{n0}	0.8 ms	0.801366 ms	0.80155 ms
t_{n1}	5.0 ms	5.012294 ms	4.99326 ms
g_h	1.21 mS/cm ²	1.22739 mS/cm ²	1.000 mS/cm ²
E_h	-40.0 mV	-39.35869 mV	-32.745 mV
vhc	-75.0 mV	-76.1190 mV	-60.9958 mV
dvhc	-11.0 mV	-11.1223 mV	-7.6066 mV
t_{hc0}	0.1 ms	0.096322 ms	0.12236 ms
t_{hc1}	193.5 ms	190.2612 ms	67.7495 ms
vhct	-80 mV	-81.0633 mV	-59.6439 mV
dvhct	21.0 mV	20.7812 mV	18.279 mV

The “strong” stimulus is 1.85 times in amplitude of the “weak” stimulus. Only the “strong” stimulus hyperpolarizes the neuron sufficiently to activate the I_h current. The “weak” stimulus only weakly activates I_h and the reduced information on the activity of I_h leads to reduced accuracy in estimating the parameters of I_h

5.2.9 NaKLh Data Presented to an NaKL Model

We contrast these results with those arising when we present the wrong data to our NaKL model. This is an example of our asking a neuron model to reproduce dynamical trajectories present in the five dimensional space of the NaKLh neuron with the four dynamical degrees of freedom in a less complex model, NaKL, of the neuron. Intuitively and mathematically, without a special circumstance in which the extra degree of freedom was somehow suppressed, this would appear impossible. It is.

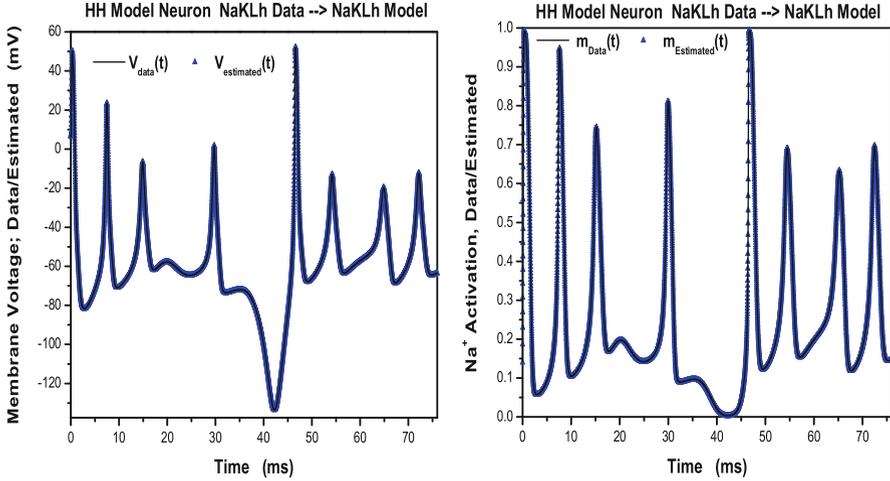


Fig. 5.7 *Left* Voltage estimated with a NaKLhHH model presented with data created from an NaKLh model. This is a twin experiment. The stimulus used is our “strong” stimulus and has $\rho_I = 0.5$ which we determined earlier allows accurate estimations. *Right* Estimation of the Na^+ activation variable $m(t)$ when NaKLh voltage data is presented to a NaKLh model. The high quality of this estimation is consistent with the $R(t)$ values in Fig. 5.8

Again we generated data using our NaKLh model with a strong stimulus and presented these data to our NaKL model. In Fig. 5.9, *Left*, we show that a rather accurate, perhaps impressive “fit” to the data is achieved by the dynamical data assimilation procedure: the variational principle with no model error. However, when we evaluate $R(t)$ for this configuration, we see the result in Fig. 5.10 which indicates the apparent success of the estimate comes from large values of the control $u(t)$ because $R(t)$ makes significant excursions to values less than one, indeed to values near zero.

In a twin experiment we are able to investigate further features of this failure by examining the ability of the model to estimate unobserved state variables. In Fig. 5.9, *Right*, we display the known and the estimated Na^+ activation variable $m(t)$ associated with the rather good representation of the observed voltage in Fig. 5.9, *Left*. The inaccuracy of the estimation of the unobserved gating variable is clear. Though we do not display it here, this inaccurate result for $m(t)$ is matched in the behavior of $h(t)$ and $n(t)$ and in the estimation of the model parameters. If we were to use the outcome of this data assimilation effort on $\text{NaKLh} \rightarrow \text{NaKL}$ we would see a failed result. On this basis we would say the NaKL model is invalid in its attempt to represent the NaKLh data. We emphasize this because outside a twin experiment, we have prediction as our only tool to establish the quality or validity of the state of the model at the end of observations and the quality of the estimates of the fixed parameters.

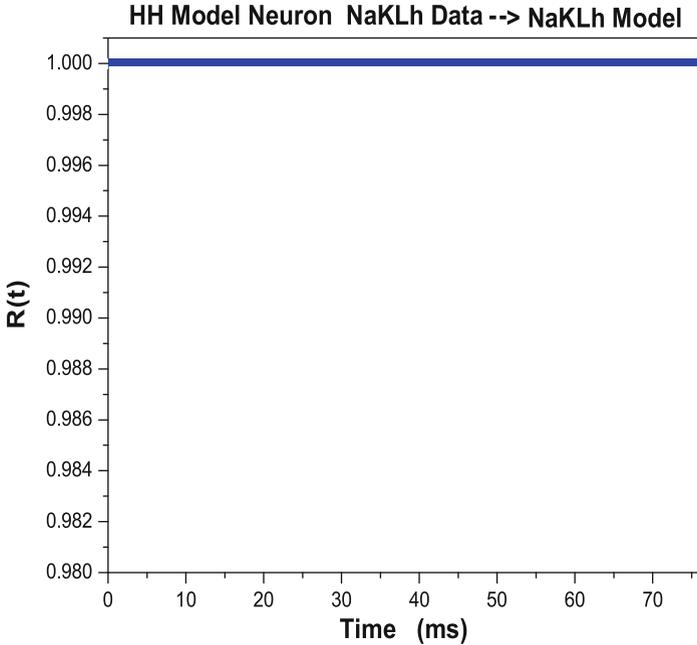


Fig. 5.8 $R(t)$ for the situation represented in Fig. 5.7. The value of $R(t)$ does not deviate from unity, indicating the model is consistent with the data

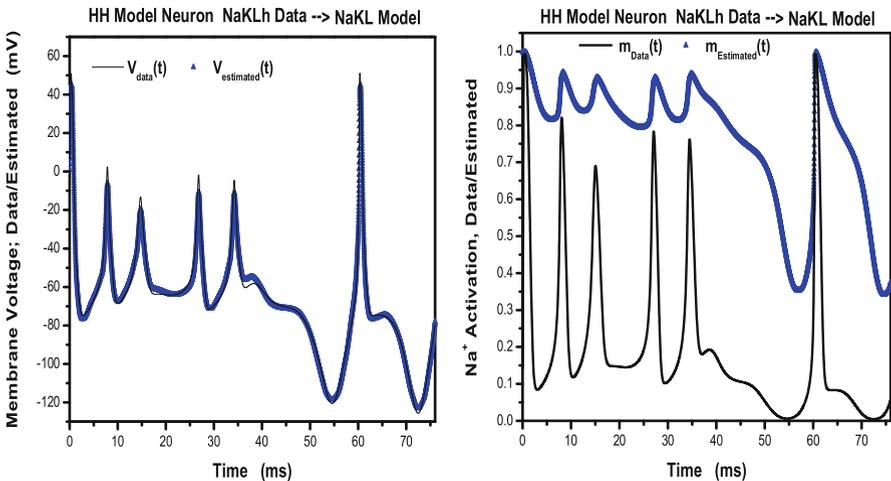


Fig. 5.9 *Left* Estimating the membrane voltage in the NaKL model when it is presented with voltage output from an NaKLh model. The accuracy of the “fit” to the voltage data is misleading as the voltage data is incorrect for this model. *Right* Estimation of the Na^+ activation variable $m(t)$ when NaKLh voltage data is presented to a NaKL model. The low quality of this estimation is consistent with the $R(t)$ value in Fig. 5.10

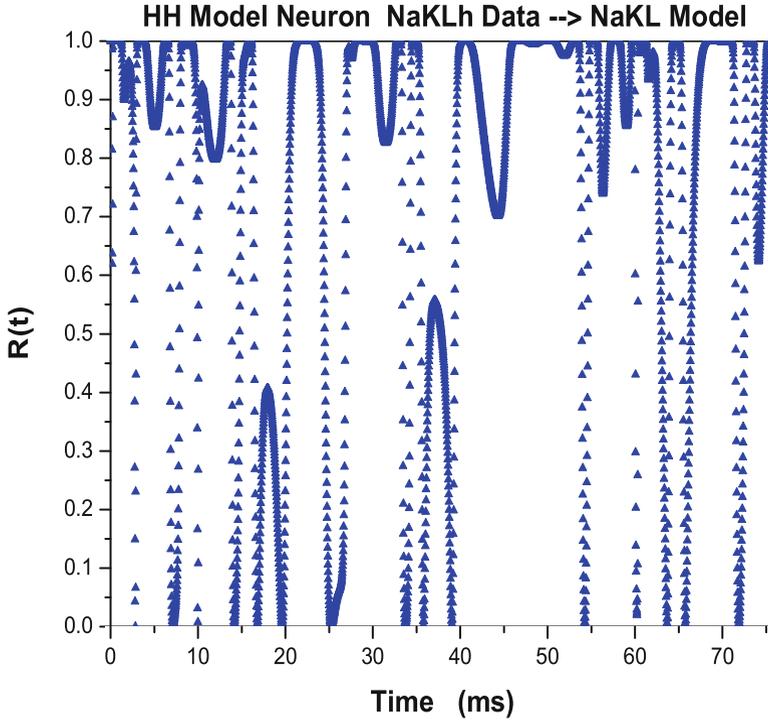


Fig. 5.10 $R(t)$ for the situation represented in Fig. 5.7. The value of $R(t)$ deviates significantly from unity, indicating the NaKL model is inconsistent with the data from a NaKLh neuron

5.2.10 *NaKL Data Driving an NaKLh Model; Pruning a Big Model*

In building a model for an experimentally observed neuron, we face the following problem: once we have selected an HH framework for the model, which currents shall we include in the model? One rather straightforward idea is to include in the model “all” the currents one can plausibly argue should enter the biophysics of that neuron. All such currents will have the form (ModelDB 2012; Graham 2002) (5.1). The kinetics of the gating variables $mc(t), hc(t)$ must be selected as well. One would expect that if any channel we have selected in this “inclusive” protocol is in fact absent in the data from the experimental preparation, then g_{ion} will be zero for that channel.

To examine whether the methods we have developed are able to identify which currents in a model are absent in the data, and thus prune a larger model down to one required by the data, we presented data from an NaKL model to a larger model having an I_h current as well. The NaKL data was generated by the NaKLh model where we eliminated I_h by selecting $g_h = 0$. Constants for the I_h kinetics were

Table 5.5 NaKL data presented to an NaKLh model

Name	Value in “data”	Estimated value
g_{Na}	120 mS/cm ²	120.4258 mS/cm ²
E_{Na}	55.0 mV	54.92950 mV
g_K	20 mS/cm ²	20.05503 mS/cm ²
E_K	-77.0 mV	-77.01366 mV
g_L	0.3 mS/cm ²	0.2990211 mS/cm ²
E_L	-54.4 mV	-54.24781 mV
$v_m = v_{mt}$	-34.0 mV	33.97316 mV
$dv_m = dv_{mt}$	34.0 mV	34.00498 mV
t_{m0}	0.01 ms	0.009585590 ms
t_{m1}	0.5 ms	0.5008398 ms
$v_h = v_{ht}$	-60.0 mV	-59.98665 mV
$dv_h = dv_{ht}$	-19.0 mV	-18.97535 mV
t_{h0}	0.2 ms	0.1998151 ms
t_{h1}	8.5 ms	8.525130 ms
$v_n = v_{nt}$	-65.0 mV	-64.94673 mV
$dv_n = dv_{nt}$	45.0 mV	45.01288 mV
t_{n0}	0.8 ms	0.7997343 ms
t_{n1}	5.0 ms	5.016256 ms
g_h	0 mS/cm ²	1.907150e-09 mS/cm ²
E_h	-40.0 mV	-60.57007 mV
v_{hc}	-75.0 mV	-6.88913 mV
dv_{hc}	-11.0 mV	4.203612 mV
t_{hc0}	0.1 ms	3.589662 ms
t_{hc1}	193.5 ms	179.5280 ms
v_{hct}	-80 mV	-44.07163 mV
dv_{hct}	21.0 mV	55.30240 mV

Parameters in a NaKLh HH model where the maximal conductance of the I_h current was set to zero, so only the NaKL model was active. The dynamical estimation procedure yielded a maximal conductance for I_h of 1.907150e-09 mS/cm². The other I_h parameters are badly estimated, but they are not relevant as the I_h current was absent in the data, though present as a possibility in the NaKLh model

specified and estimated although the I_h contribution to the voltage equation was absent. Table 5.5 displays the results for this situation.

We see that all the information about the Na⁺, K⁺, and leak currents was accurately reproduced, the absence of the I_h current was recorded as a very small g_h , namely, $g_h = 1.9 \times 10^{-9}$ mS/cm², and the estimates for E_h and the parameters in the kinetics for I_h are wrong. Because the h-current is, in fact, absent from the data, though present in the model, the inaccuracy of an estimation of its parameters is of no physical consequence. The idea that we could build a “large” model of all neurons and use the data to prune off currents that are absent is both made plausible and supported by this calculation.

5.3 The Lorenz96 Model

We have explored examples from electrical circuits, fluid dynamics, and neurobiology in various parts of this book. We now turn to an example which is widely used in the geophysical literature to illustrate methods of statistical data assimilation. It is a collection of ordinary differential equations first studied by Lorenz (1996) with state variables set out on a ring, presumably representing equatorial locations. After describing the model we examine it in the case of a variational principle data assimilation with no model error, then we turn to Monte Carlo evaluation of the full data assimilation path integrals. This example is illustrative both because the solutions of the dynamical equations are chaotic but also because with some ease we are able to study high dimensional versions of the equations.

From the rather simple set of Lorenz96 equations, we then turn to a widely used component of numerical weather prediction models called the shallow water equations. In each example we find that many measurements are required to allow searches over fixed parameters and state variables to permit accurate prediction.

In a discussion of predictability in atmosphere and ocean models, Lorenz (1996) introduced a model with K state variables located on a ring. The model is meant to illustrate the forced westward flow of dynamical information transported by an “advection” quadratic in the dynamical variables. It was subsequently used by Lorenz and Emanuel (1998) to discuss strategies in selecting additional locations for dynamical weather measurements for the greatest improvement in analysis and forecasting. They asked, for a given number K of observation stations, whether predictions would be improved by having $K + 1$ stations.

The model has one fixed parameter, the “forcing” f , and K dynamical variables $x_a(t)$ $a = 1, 2, \dots, K$ on a ring. The dynamical equations are

$$\frac{dx_a(t)}{dt} = x_{a-1}(t)(x_{a+1}(t) - x_{a-2}(t)) - x_a(t) + f, \quad a = 1, 2, \dots, K \quad (5.11)$$

with $x_{-1}(t) = x_{K-1}(t)$, $x_0(t) = x_K(t)$, and $x_{K+1}(t) = x_1(t)$.

A fixed point solution of these equations is $x_a(t) = f$, and small perturbations about this solution, $x_a(t) = f + \delta x_a(t)$ yield in linear order in $\delta x_a(t)$

$$\frac{d\delta x_a(t)}{dt} = f(\delta x_{a+1}(t) - \delta x_{a-2}(t)) - \delta x_a(t). \quad (5.12)$$

If we write $\delta x_a(t) = A_k(t)e^{ika}$, then the amplitudes $A_a(t)$ satisfy

$$\frac{dA_k(t)}{dt} = A_k(t) \left[(e^{ik} - e^{-2ik})f - 1 \right], \quad (5.13)$$

and when the real part of the quantity in square brackets is larger than unity, the fixed point is unstable. The maximum of $\cos k - \cos 2k$ is $9/8$ for $0 \leq k \leq 2\pi$, so when $f > 8/9$ the fixed point is unstable. The solution then proceeds through a sequence

of bifurcations including limit cycle (periodic) solutions, and when $f \approx 6.5$ chaotic solutions appear.

It is our plan to use this model, with various values of K in a twin experiment setting. We will hold the forcing fixed for much of our discussion and use a variational principle with no model error to estimate the unobserved states of the model at the end of an observation window, $t_m = T$. Using the known parameters and the estimated state $\mathbf{x}(T)$, we will predict for $t > T$. The orbits of the model with the forcing we select will be chaotic, so the prediction horizon will be limited by this global instability. After this procedure we will turn to a Monte Carlo evaluation of the path integral for this model.

5.3.1 Synchronization

We start with $K = 5$ and generate twin experiment “data” $y_a(t)$; $a = 1, 2, \dots, 5$ using a time step of $\Delta t = 0.01$, which corresponds to about $\frac{6}{5}$ h in the scales defined by the original model (Lorenz and Emanuel 1998). Orbits generated with these parameters are chaotic. Using the usual nonlinear time series analysis (Abarbanel 1996; Kantz and Schreiber 2004) for one of the signals from this data set, we determined that the Lyapunov dimension associated with these exponents is $D_L = 4.6$.

Recall that we argued that proper “tuning” of the model as a receiver of transmissions from data as a transmitter requires synchronization of the model output corresponding to the measurements; $y_l(n) \approx h_l(\mathbf{x}(n))$.

We now ask if we are able to synchronize the model Eq. (5.11) for $x_a(t)$ with the same forcing parameter using information on $y_1(t)$ alone. To address this we required several calculations. We first coupled $y_1(t)$ into the equation for $x_1(t)$ in the following manner:

$$\begin{aligned}
 \frac{dx_1(t)}{dt} &= x_5(t)(x_2(t) - x_4(t)) - x_1(t) + f + k_1(y_1(t) - x_1(t)) \\
 \frac{dx_2(t)}{dt} &= x_1(t)(x_3(t) - x_5(t)) - x_2(t) + f \\
 \frac{dx_3(t)}{dt} &= x_2(t)(x_4(t) - x_1(t)) - x_3(t) + f \\
 \frac{dx_4(t)}{dt} &= x_3(t)(x_5(t) - x_2(t)) - x_4(t) + f \\
 \frac{dx_5(t)}{dt} &= x_4(t)(x_1(t) - x_3(t)) - x_5(t) + f.
 \end{aligned} \tag{5.14}$$

We found that this did not result in synchronization, and we elaborate on this in a moment.

We then coupled both $y_1(t)$ and $y_3(t)$ into the model equations for $x_a(t)$

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= x_5(t)(x_2(t) - x_4(t)) - x_1(t) + f + k_1(y_1(t) - x_1(t)) \\
\frac{dx_2(t)}{dt} &= x_1(t)(x_3(t) - x_5(t)) - x_2(t) + f \\
\frac{dx_3(t)}{dt} &= x_2(t)(x_4(t) - x_1(t)) - x_3(t) + f + k_2(y_3(t) - x_3(t)) \\
\frac{dx_4(t)}{dt} &= x_3(t)(x_5(t) - x_2(t)) - x_4(t) + f \\
\frac{dx_5(t)}{dt} &= x_4(t)(x_1(t) - x_3(t)) - x_5(t) + f,
\end{aligned} \tag{5.15}$$

introducing information into the model dynamics beyond that carried by $y_1(t)$ alone. In this case we find for either $k_1 = 0$ and $k_2 \neq 0$ or $k_2 = 0$ and $k_1 \neq 0$, there is no synchronization of the data and the model. This is equivalent to the statement with one coupling only. However, when both k_1 and $k_2 \approx 3$, the data and the model synchronize.

The synchronization error given by

$$C(k_1, k_2) = \frac{1}{2N} \sum_{n=0}^{m-1} \left\{ (y_1(t_n) - x_1(t_n))^2 + (y_3(t_n) - x_3(t_n))^2 \right\}, \tag{5.16}$$

is evaluated by solving Eq. (5.15) for initial conditions $\{x_a(0)\}$ which in general are different than those in the data $\{y_a(0)\}$, and starting the sum (5.15) at t_0 after transients are discarded. In Fig. 5.11 we show a cut through the three dimensional surface $\{k_1, k_2, C(k_1, k_2)\}$ with $k_1 = k_2 = k$ in the range $[0, 20]$. For $k \approx 3$ we see that the synchronization error has gone to zero.

More information is in the full three-dimensional surface $\{k_1, k_2, C(k_1, k_2)\}$ shown in Fig. 5.12. Here we can see that when either k_1 or k_2 is 0, so only one piece of information is coupled into the Lorenz96 model for $x_a(t)$, synchronization does not occur. When both k_1 and k_2 are larger than approximately three, synchronization is achieved.

We have indicated throughout this book that synchronization is important for an accurate estimation of the parameters and state variables of the model system. We can see that in operation in our present example as well. We generated the solution to the differential equation (5.15) with a value of $x_a(0)$ different from that $y_a(0)$ used in producing the data. If we wished to find one of the $y_a(0)$, we would expect that it may be estimated by searching for a minimum of the synchronization error as a function of $x_2(0)$. We evaluated the cost function

$$C(k_1, k_2, x_2(0)) = \frac{1}{2N} \sum_{n=0}^{m-1} \left\{ (y_1(t_n) - x_1(t_n))^2 + (y_3(t_n) - x_3(t_n))^2 \right\}, \tag{5.17}$$

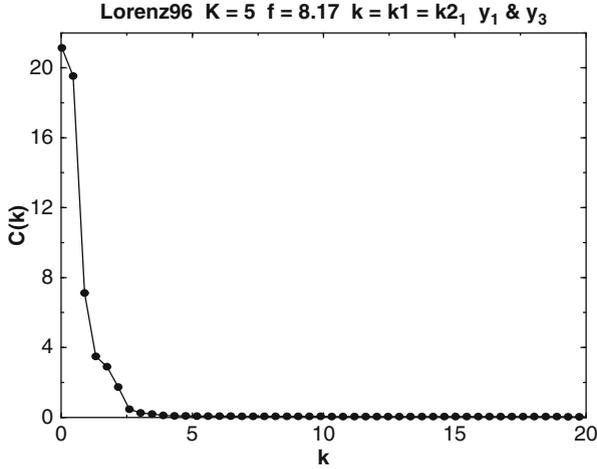


Fig. 5.11 Lorenz96 model; $K = 5$. We display the synchronization error $C(k_1, k_2) = \frac{1}{2N} \sum_{n=0}^{m-1} \left\{ (y_1(t_n) - x_1(t_n))^2 + (y_3(t_n) - x_3(t_n))^2 \right\}$ as a function of k with $k_1 = k_2 = k$ associated with coupling information from $y_1(t)$ and $y_3(t)$ into the equations for $x_1(t)$ and $x_3(t)$ as in Eq. (5.15). The two systems now synchronize for k larger than about 3

for four different selections of k_1, k_2 as a function of the choice of $x_2(0)$. Our data was generated using $y_2(0) \approx 4.9$. The variation of $C(k_1, k_2, x_2(0))$ in $x_2(0)$ for the sets of values $\{k_1, k_2\} = \{(0.0, 0.0), (8.0, 0.0), (0.0, 8.0), (8.0, 8.0)\}$ is shown in Fig. 5.13. In the Upper Left Panel $\{k_1 = 0.0, k_2 = 0.0\}$; in the Upper Right Panel $\{k_1 = 8.0, k_2 = 0.0\}$; and in the Bottom Left Panel $\{k_1 = 0.0, k_2 = 8.0\}$. We see that for these choices of coupling, where the couplings can be zero, the variation in $x_2(0)$ is full of local minima which have little to do with the known value $y_2(0) \approx 4.9$.

In the Lower Right Panel, however, when $\{k_1 = 8.0, k_2 = 8.0\}$ and the model output $x_a(t)$ and the data $y_a(t)$ are synchronized, the variation in $x_2(0)$ of the cost function has become smooth with an easy to locate minimum near the correct value. We have seen this kind of behavior in earlier examples, and we repeat it here as this is the first time we have required more than one dynamical piece of information to be passed from the data to the model to achieve synchronization.

Later we will examine a larger model, a small grid approximation to geophysical partial differential equations, and demonstrate that lack of synchronization is directly connected to lack of predictability. The unobserved state variables are inaccurately estimated leading to inaccurate state values from which prediction proceeds.

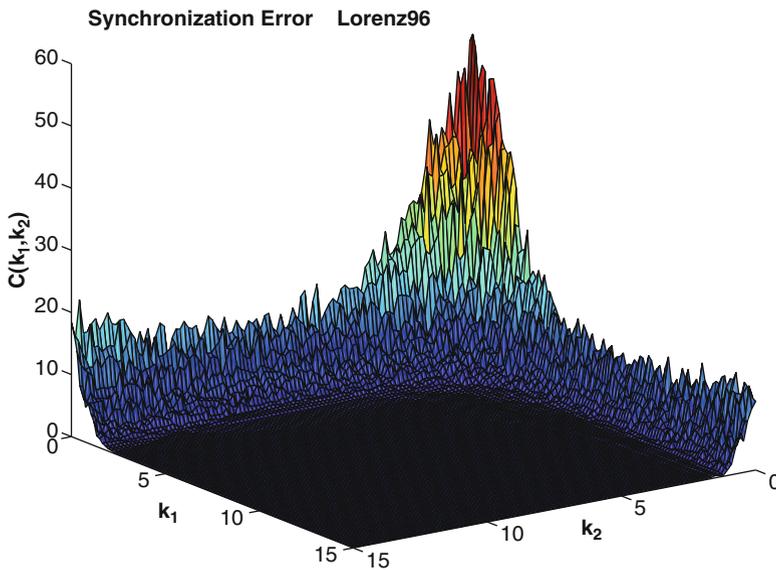


Fig. 5.12 Lorenz96 model; $K = 5$. We display the synchronization error $C(k_1, k_2)$ as a function of k_1, k_2 associated with coupling information from $y_1(t)$ and $y_3(t)$ into the equations for $x_1(t)$ and $x_3(t)$ as in Eq. (5.15). The two systems now synchronize for k_1, k_2 each larger than about 3. This is an indication that two unstable directions are present on the synchronization manifold $\mathbf{x}(t) = \mathbf{y}(t)$. Figure 5.11 is a cut along the $k_1 = k_2$ diagonal of this graphic

5.3.2 Lyapunov Exponents

We begin with a repeat of some material from Sect. 2.2.3, and then we enlarge on it. To evaluate the Lyapunov exponents for a differential equations such as the Lorenz96 model, we analyze perturbations about an orbit of the dynamical system (Abarbanel 1996; Kantz and Schreiber 2004). The orbit is generated by solving the differential equation for the Lorenz96 model (5.11) numerically from a selected $\mathbf{x}(0)$ by discretizing time and providing a rule guiding movement from the state at time t_n ; $\mathbf{x}(n)$ to $\mathbf{x}(n + 1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$. The stability of orbits when the K -dimensional Lorenz96 model operates autonomously, that is, without external forcing or information, is determined by the K eigenvalues e^{λ_a} of the Oseledec matrix Eq. (2.28). If any of the K values of λ_a is positive, the solution is chaotic. One eigenvalue is always zero as we are discussing a flow in continuous time.

In Fig. 5.14 we show the computed values of the λ_a for $K = 5$ as a function of the forcing strength f . The zero exponent is not shown. We have selected $f = 8.17$ and see that one Lyapunov exponent is positive leading to chaotic solutions of the model equations.

When we wish to pass information from observations $\mathbf{y}(t_n)$ to the model, one way to do it is by coupling them to the model via

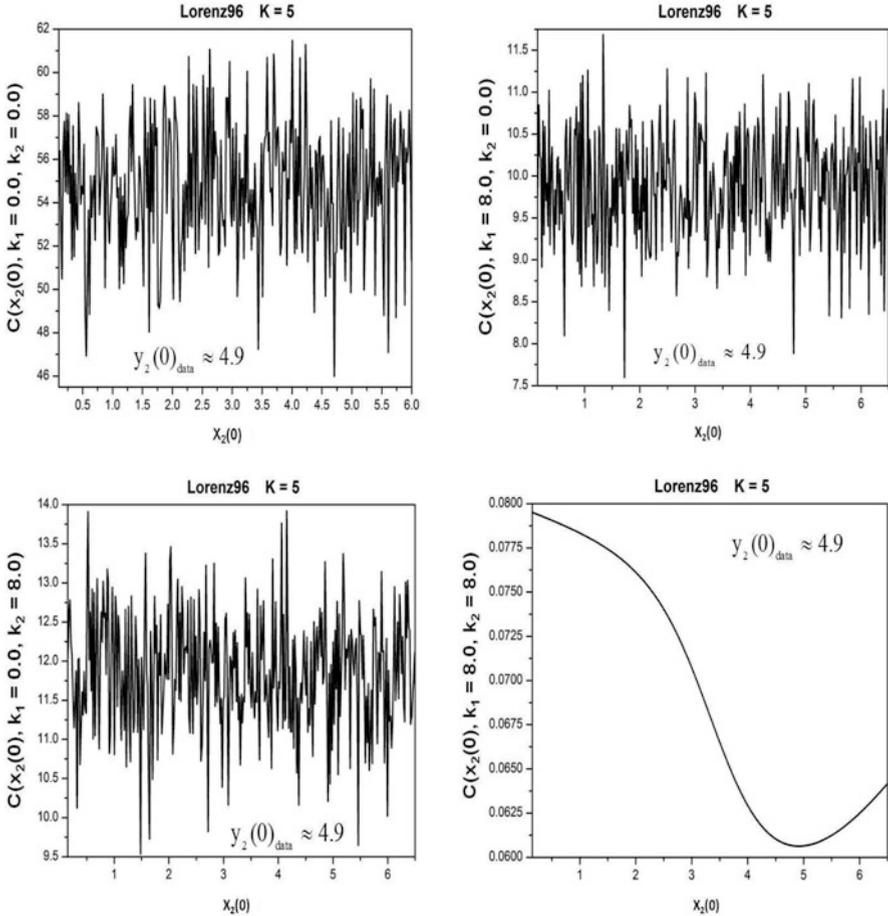


Fig. 5.13 Cross sections through the cost function $C(k_1, k_2, x_2(0))$ for the Lorenz96, $K = 5$ model through measurements of $y_1(n)$ and $y_3(n)$. We display the dependence of this cost function, Eq. (5.15) on the initial value $x_2(0)$ of one of the model state variables. We examine four combinations of the couplings $\{k_1, k_2\} = \{(0.0, 0.0), (8.0, 0.0), (0.0, 8.0), (8.0, 8.0)\}$. For the first three of these, the model and the data are not synchronized. For the final value, the model and the data synchronize. In the last case, we see that the search over values of $x_2(0)$ for the value of this initial state variable in the data, namely $y_2(0) \approx 4.9$ can be successful. In the unsynchronized cases, such a search would be significantly impeded by the multiple local minima

$$x_a(n+1) = f_a(\mathbf{x}(n)) + \sum_{b=1}^K Q_{ab}(y_b(n) - x_b(n)); \quad a, b = 1, 2, \dots, K, \quad (5.18)$$

and the $K \times K$ coupling matrix Q_{ab} has nonzero elements only where data is coupled into the model. In the Oseledec result, one requires the Jacobian of the dynamics iterated along the chaotic orbit, and with the coupling, the Jacobian is

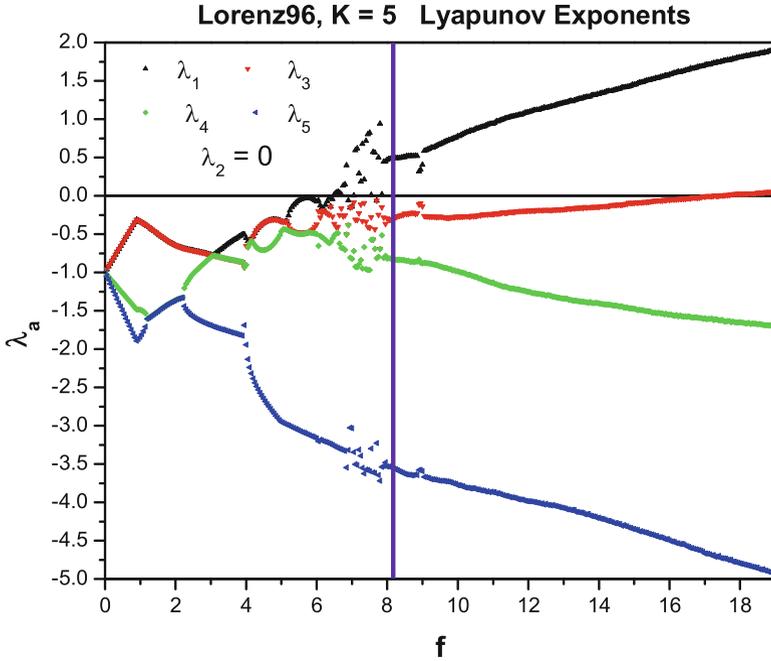


Fig. 5.14 Nonzero Lyapunov exponents for the Lorenz96, $K = 5$, model as a function of the forcing f in the model. One Lyapunov exponent is always 0 (Abarbanel 1996), and it is not displayed. The largest Lyapunov exponent becomes positive around $f \approx 6.5$, and for larger values of f the system is chaotic. In the neighborhood of $f \approx 18$ a second positive Lyapunov exponent appears. We selected $f = 8.17$, shown by the vertical violet line, for our discussion in the main text (Kostuk 2012)

$$\frac{\partial f_a(\mathbf{x}, p)}{\partial x_b} - Q_{ab}. \tag{5.19}$$

If all the conditional Lyapunov exponents (CLEs) of this quantity, evaluated via the Oseledec matrix, are negative, then the solutions of Eq. (5.18) are stable. There is no requirement that one among the collection of CLEs need be zero, unless $Q_{ab} = 0$.

We have evaluated the CLEs (Kostuk 2012) for the case $Q_{ab} = k\delta_{a1}\delta_{b1}$ and $f = 8.17$, and we display the result in Fig. 5.15 as a function of k . As we expect, at $k = 0$ one of the exponents is zero, and one is positive. This corresponds to the values of the λ_a at the vertical violet line in Fig. 5.14. As k increases, four of the CLEs become negative, while the largest one decreases but remains positive. This tells us that providing information from the data to the Lorenz96 model from a single data stream does not stabilize the synchronization manifold. This is consistent with the message in Fig. 5.13. Indeed, as we have noted, the origin of the multiple local minima in this figure is the incoherence of two chaotic time series: one from the data and one from the model output.

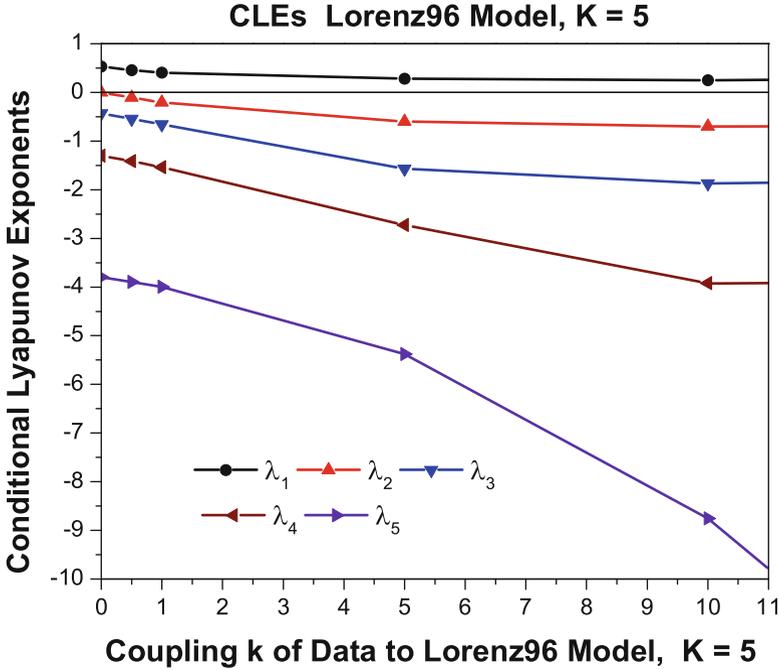


Fig. 5.15 Conditional Lyapunov exponents (CLEs) for the Lorenz96 model, $K = 5$, as a function of the constant coupling $k(y_1(t) - x_1(t))$ in the dynamical equation for $x_1(t)$. One CLE remains positive indicating that the synchronization manifold on which $y_1(t) = x_1(t)$ is unstable. At $k = 0$ one Lyapunov exponent is positive indicating the dynamical system yield periodic orbits. At $k = 0$ one Lyapunov exponent is 0, as we expect from a system of differential equations (Abarbanel 1996; Kostuk 2012)

Taking a lesson from the Lower Right panel of Fig. 5.13, we now select

$$Q_{ab} = k_1\delta_{a1}\delta_{b1} + k_2\delta_{a3}\delta_{b3}, \tag{5.20}$$

and evaluate the CLEs as a function of $k = k_1 = k_2$. In Fig. 5.16 we now see that at $k = 0$, we agree with the previous calculation as we must. As k increases, all five CLEs decrease and between $k = 2$ and $k = 3$ the largest CLE moves from a positive value to a negative value. The data sources $y_1(t)$ and $y_3(t)$ are sufficient to stabilize the synchronization of the data and the model. In the Lower Right panel of Fig. 5.13, we see that as a consequence we are now able to accurately estimate the state variable $x_2(0)$.

The synchronization of the data and the model should allow the estimation of the three unobserved state values $x_2(t), x_4(t), x_5(t)$, and we will show this to be so. The discussion of such situations has been a long-standing topic in the control theory literature wherein it is called observability (Nijmeijer and Mareels 1997). In defining observability one usually require synchronization in the form

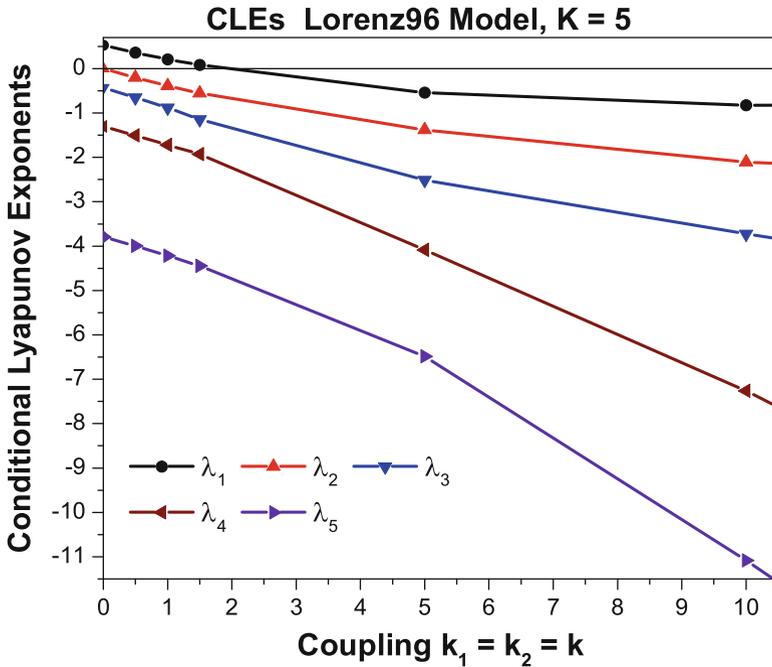


Fig. 5.16 Conditional Lyapunov exponents (CLEs) for the Lorenz96 model, $K = 5$, as a function of the constant couplings $k_1(y_1(t) - x_1(t))$ and $k_2(y_3(t) - x_3(t))$ in the dynamical equations for $x_1(t)$ and for $x_3(t)$. The largest CLE becomes negative near $k \approx 3$ indicating that the synchronization manifold on which $y_1(t) = x_1(t)$; $y_3(t) = x_3(t)$ is stable. At $k = 0$ one Lyapunov exponent is positive indicating the dynamical system yield periodic orbits. At $k = 0$ one Lyapunov exponent is 0, as we expect from a system of differential equations (Abarbanel 1996). This consistent with the information in Figs. 5.11 and 5.12 (Kostuk 2012)

$|\mathbf{x}(t) - \mathbf{y}(t)| \rightarrow 0$ as $t \rightarrow \infty$, which is a stronger form than our synchronization criterion which is an average over segments of the orbits $\mathbf{x}(t)$ and $\mathbf{y}(t)$. In practice, these do not appear to be different requirements. One can safely say that the notion of observability, which goes back to the 1960s (Luenberger 1964, 1966, 1971, 1979; Nijmeijer and Mareels 1997), is, in essence, the same as the stability of the synchronization manifold.

The calculation of the CLEs, or the Lyapunov exponents when $k_j = 0$, requires the determination of the eigenvalues of a product of Jacobian matrices evaluated along an orbit of the dynamical system. This product enters the Oseledec matrix and is a very ill-conditioned $D \times D$ matrix. If the eigenvalues of the matrix are ordered so $\lambda_1 > \lambda_2 > \dots > \lambda_D$, the condition number of the matrix is $e^{2m(\lambda_1 - \lambda_D)}$ for m steps along a trajectory. This can be a very large number, and diagonalizing such a matrix requires substantial care. We use the recursive QR decomposition method described in Abarbanel (1996) which is stable and accurate. It also takes a long time to run, and as D increases it may not be an attractive procedure.

5.3.3 CLEs for Lorenz96 Models

How can we know the number of required measurements to assure that the largest CLE for the Lorenz96 model with D dimensions is negative? A direct way is to find the eigenvalues of the Oseledec matrix (Oseledec 1968) for Eq. (5.18). This requires the diagonalization of the product of $D \times D$ Jacobian matrices.

Another idea is to move onto the synchronization manifold (Kostuk 2012) where L measured data time series $y_l(t)$; $l = 1, 2, \dots, L$ are synchronized to functions $h_l(\mathbf{x}(t))$ and examine the Oseledec matrix for the problem where L components of the model state $\mathbf{x}(t)$ are replaced by the $\mathbf{y}(t)$ (Kostuk 2012). We adopt our usual simplification $h_l(\mathbf{x}) = x_l(t)$.

In the case $L = 1$, the dynamics is

$$x_k(n+1) = f_k(y_1(n), x_k(n), \mathbf{p}); \quad k = 2, 3, \dots, D, \quad (5.21)$$

and with $y_1(n)$ and $y_2(n)$ passed to the dynamical rule, we require the Oseledec matrix for

$$x_k(n+1) = f_k(y_1(n), y_2(n), x_k(n), \mathbf{p}); \quad k = 3, \dots, D, \quad (5.22)$$

and so forth. If we analyze these equations for given $y_l(n)$, we determine $D - L$ exponents using the Oseledec matrix as in the case of autonomous equations. These exponents are CLEs as they are conditioned on the given $\mathbf{y}(n)$. $D - L$ CLEs will be evaluated determining the stability of the synchronization manifold $x_l(n) = y_l(n)$ for each L . In a formal sense, this procedure is equivalent to setting the couplings $k \rightarrow \infty$. It is also the synchronization scheme first explored by Pecora and Carroll (1990).

CLE calculations for various D and L for Lorenz96 models are now presented. The models are integrated for 10^5 time points; these initial values are discarded. The Lyapunov exponents and CLEs are then evaluated with times series from subsequent trajectories of 10^9 points. All integration is done using a fourth-order Runge-Kutta integration with a time step of $\Delta t = 0.01$. The Jacobians are numerically estimated using the trajectory results based upon an averaged offset in state space $\pm 10^{-8}$ for all directions.

We begin with $D = 5$. In Table 5.6 we report the results of this calculation for the Lyapunov exponents when $L = 0$, the autonomous model differential equations, and the CLEs for $L = 1$ and $L = 2$. Other $L = 2$ coupling schemes are the same as those reported by cyclic symmetry of the $D = 5$ model.

Next we move to the case $D = 9$ for the Lorenz96 equations. In Table 5.7 we have listed both the Lyapunov exponents of the autonomous $D = 9$ model, as well as the CLEs that result from an $L = 1$ coupling. With a single coupling there always remains a positive CLE, indicating that the synchronization manifold remains unstable.

Table 5.6 The Lyapunov exponents of the autonomous Lorenz96, $D = 5$ model ($L = 0$) and the conditional Lyapunov exponents of the Lorenz96 $D = 5$ model with $f = 8.17$ using various coupling schemes; $L = 1, 2$

CLEs for Lorenz96 model $D = 5$ $L = 0, 1, 2$

	$L = 1$	$L = 2$	$L = 2$
$L = 0$	y_1	y_1, y_2	y_1, y_3
0.527	0.491	-0.117	-0.106
10^{-7}	-0.298	-1.201	-1.089
-0.432	-0.832	-1.871	-2.183
-1.303	-3.550	-	-
-3.793	-	-	-

Both $L = 2$ coupling schemes result in stabilized synchronization manifolds (Kostuk 2012)

Table 5.7 The Lyapunov exponents ($L = 0$) and singly coupled CLEs ($L = 1$) for the Lorenz96 $D = 9$ model with $f = 8.17$

CLEs for Lorenz96 model
 $D = 9$ $L = 0, 1$

$L = 0$	$L = 1$
1.268	1.012
0.585	0.447
10^{-6}	-0.092
-0.208	-0.628
-0.759	-1.125
-1.245	-1.681
-1.815	-2.459
-2.573	-3.632
-4.252	-

The single coupling term reduces the positive Lyapunov exponents; however, it is insufficient to provide stable motion on the synchronization manifold (Kostuk 2012)

We next select $L = 2$. From Table 5.8, we see that *none* of the coupling schemes that incorporate two observations into the Lorenz96 $D = 9$ model will result in synchronization. It is not until three observations are used that this occurs. The CLEs that result from some of the $L = 3$ possible coupling schemes are listed in Table 5.9. We have left out some of the coupling possibilities that are redundant based upon the cyclic symmetry of this model.

As seen in Table 5.9 some coupling schemes are more successful than others at reducing the any positive CLEs to negative values. For example the $y_1, y_2,$ and y_3 arrangement does not lead to a completely negative spectrum of CLEs, whereas the $y_1, y_2,$ and y_7 arrangement does accomplish this.

Table 5.9 displays the CLEs arising when $L = 3$ data streams are presented to the Lorenz96 $D = 9$ model. When all CLEs are negative, the synchronization manifold is stable. This is not the case for all possible coupling of three data streams into the model. This is consistent with our formulation of the general problem of statistical data assimilation where the information passed to the model depends on the observation and on the model.

Table 5.8 The CLEs of the Lorenz96 $D = 9$ model with two coupled observations

CLEs for Lorenz96 model $D = 9 \quad L = 2$			
y_1, y_2	y_1, y_3	y_1, y_4	y_1, y_5
0.162	0.789	0.638	0.799
-0.208	0.189	0.058	0.265
-0.581	-0.347	-0.458	-0.380
-0.909	-0.880	-0.953	-0.930
-1.281	-1.425	-1.449	-1.511
-1.794	-2.225	-2.087	$v - 2.342$
-2.543	-3.414	-3.061	$v - 3.213$

Each column represents a different coupling scheme with two pieces of information $y_1(t)$ and $y_i(t)$, for $i = 2, 3, 4, 5$. The coupling schemes not listed are equivalent to one that is listed via symmetry. None of these coupling schemes with only two data streams is capable of synchronizing the $D = 9$ model output with the data (Kostuk 2012)

Table 5.9 The CLEs of the Lorenz96 $D = 9$ model with three coupled observations

CLEs for Lorenz96 model $D = 9 \quad L = 3$					
y_1, y_2, y_3	y_1, y_2, y_7	y_1, y_2, y_8	y_1, y_3, y_5	y_1, y_3, y_7	y_1, y_4, y_7
0.106	-0.070	-0.024	0.661	0.501	0.025
-0.308	-0.553	-0.471	0.013	-0.076	-0.421
-0.736	-0.837	-0.846	-0.618	-0.732	-0.986
-1.130	-1.207	-1.137	-1.226	-1.265	-1.257
-1.653	-1.517	-1.523	-2.083	-1.991	-1.625
-2.436	-2.128	-2.311	-3.216	-2.906	-2.202

Each column represents a different coupling scheme. The choice of observed variables is important. Note that presenting y_1, y_2, y_7 or y_1, y_2, y_8 permit synchronization. Other choices do not succeed in producing negative CLEs (Kostuk 2012)

As we increase D , the number of dimensions of the model, we increase the potential for finding more positive Lyapunov exponents for the autonomous dynamics and for finding more positive CLEs as well.

According to the results we have shown we would conclude that the $D = 5$ Lorenz96 model requires two observations in order to achieve synchronization, and the $D = 9$ model requires three. We have also noted that not all coupling schemes of L data streams for a $D > L$ -dimensional model will lead to stabilization of the synchronization manifold. We have examined the minimum number of observables which we call the critical L , or L_c , required for Lorenz96 models with D as large as 45. Table 5.10 shows the results. In an approximate sense $L_c \approx 0.4D$, but this does not mean all sets of L_c observables will lead to synchronization. Not all possible coupling schemes for large D were examined as this quickly becomes very computationally intensive for large D . One requires the construction of the product of many $(D - L) \times (D - L)$ matrices and the determination of the largest CLE for each selection of L . As D becomes large and L grows, there are many such selections and most must be examined.

Table 5.10 The minimum number of observations L_c of the D -dimensional Lorenz96 model with forcing strength $f = 8.17$ required so that all the CLEs are negative $L_c \approx 0.4D$

D	4	5	9	10	11	13	15	20	45
L_c	2	2	3	4	5	6	6	8	21

This ensures synchronization of the model output with the observed data (Kostuk 2012)

5.3.4 Lorenz96 Model: Variational Principle; No Model Errors

We seek now to use this knowledge on the stability of the synchronization manifold to estimate the unobserved state variables in the Lorenz96 model for $K = 5$ and predict beyond the time window of observations. To this end we present information from two observed time series for $y_1(t)$ and $y_3(t)$ to the five dimensional dynamics. As earlier, we replace the constant values of the couplings with $k_1 \rightarrow u_1(t)$ and $k_2 \rightarrow u_2(t)$ recognizing that the strength of the coupling to stabilize the synchronization manifold may vary as we progress around the orbits of the system.

The dynamics of the $D = 5$ Lorenz96 model is now given as

$$\begin{aligned}
 \frac{dx_1(t)}{dt} &= x_5(t)(x_2(t) - x_4(t)) - x_1(t) + f + u_1(t)(y_1(t) - x_1(t)) \\
 \frac{dx_2(t)}{dt} &= x_1(t)(x_3(t) - x_5(t)) - x_2(t) + f \\
 \frac{dx_3(t)}{dt} &= x_2(t)(x_4(t) - x_1(t)) - x_3(t) + f + u_2(t)(y_3(t) - x_3(t)) \\
 \frac{dx_4(t)}{dt} &= x_3(t)(x_5(t) - x_2(t)) - x_4(t) + f \\
 \frac{dx_5(t)}{dt} &= x_4(t)(x_1(t) - x_3(t)) - x_5(t) + f,
 \end{aligned} \tag{5.23}$$

and we recall we wish to have no trace of the coupling terms $u_1(t)$ and $u_2(t)$ when our estimation protocol is completed. They regularize the search for state and parameters values on the synchronization manifold, but there is no physics associated with their value.

Adding a cost for the controls $\{u_1(t), u_2(t)\}$, the cost function takes the form

$$C(\mathbf{x}, \mathbf{u}) = \frac{1}{2m} \sum_{n=0}^{m-1} \left\{ (y_1(t_n) - x_1(t_n))^2 + (y_3(t_n) - x_3(t_n))^2 + u_1(n)^2 + u_2(n)^2 \right\}. \tag{5.24}$$

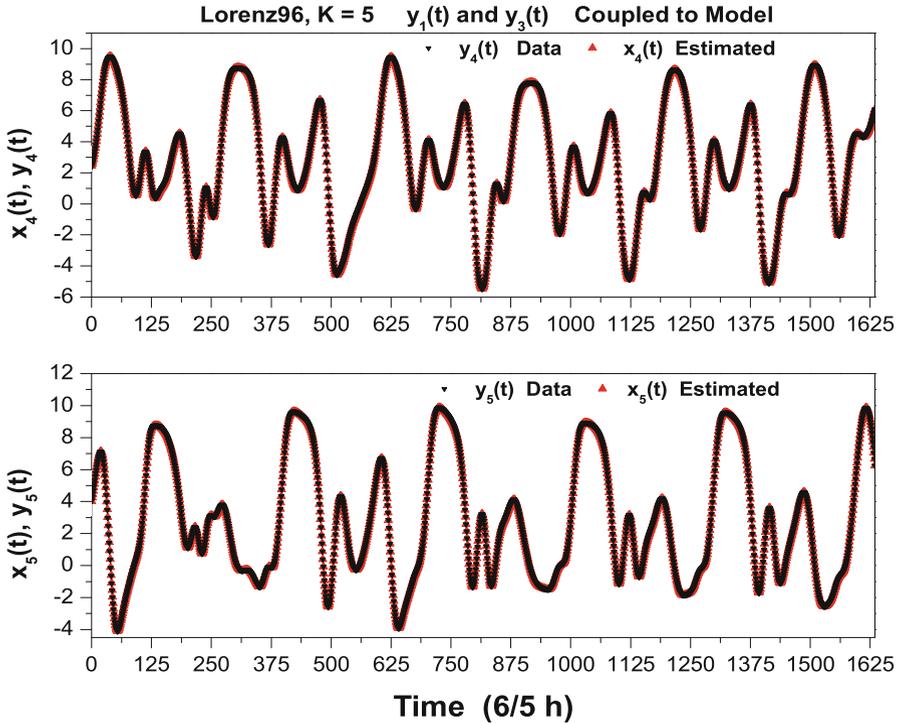


Fig. 5.17 Lorenz96 (Lorenz 1996) model. The data $y_4(t)$, $y_5(t)$ (black down triangles) and the estimated model state variables $x_4(t)$, $x_5(t)$ (red up triangles) are displayed. Both $x_4(t)$ and $x_5(t)$ are unobserved state variables

We now minimize this cost function using the direct method implemented via IPOPT as discussed in earlier chapters. We found that we could also estimate the parameter f quite accurately, but this did not change any essential part of the calculations. We proceed with $f = 8.17$.

We generated a data set from the Lorenz96 model with $K = 5$ and $f = 8.17$. This produced a chaotic time series on an attractor of dimension about 4.6. Using $m = 1,635$ data points from $y_1(t_n)$ and $y_3(t_n)$ we minimized Eq. (5.24) subject to the dynamical Equation (5.23). This produced seven time series over the interval $[t_0 = 0, t_m = T = 1,635]$: five model outputs $x_a(t_n)$ and two sequences $u_1(t_n)$, $u_2(t_n)$.

In Fig. 5.17 we see that the unobserved state variables $x_4(t)$ and $x_5(t)$ are well estimated over the observation window. We do not display the other estimated state variables nor do we show values for the $u_l(t_n)$. The quality of the estimations of the other $x_a(t)$ is equally as good as the values shown, and the actual values of the $u_l(t)$ are small. The analogue of the dimensionless consistency check $R_l(t)$; $l = 1, 2$ shows the R_l values to always be close to unity.

We now took the values of the model state variables at t_n ; $n = 1, 600$, so a few steps back into the observation region, and selecting these as initial conditions predicted forward in time for $t > 1, 600$. The choice of $t = 1, 600$ rather than the actual end of the observation window at $t = 1, 635$ was simply to provide a bit of variety in our protocols. The outcome of this prediction for the observed state variable $x_1(t)$ and the unobserved state variable $x_2(t)$ is shown in Fig. 5.18 where, for each state variable, it is compared to the known data $y_1(t)$ and $y_2(t)$. The accuracy of the estimations is supported by the accuracy of the predictions as one moves forward from the initial state at $t = 1, 600$. Because the orbits of the Lorenz96 model are chaotic, we expect that errors at $t_0 = 1, 600$ will grow in time as $\text{error}(t) \approx \text{error}(t_0)e^{\lambda_1(t-t_0)}$. Using the calculated value of the largest Lyapunov exponent $\lambda_1 \approx 0.5$ and the computed error from state estimates at $t = 1, 600$ where prediction begins, we find consistency. The same features are apparent in the prediction for $t > 1, 600$ for the two unobserved state variables $x_4(t)$ and $x_5(t)$ as shown in Fig. 5.19.

It might well go without saying, but we say it nonetheless: predicting using estimates from minimizing the Lorenz96 $D = 5$ cost function with only one allowed nonzero value of the controls $u_1(t), u_2(t)$ results in quite inaccurate predictions.

5.4 Monte Carlo Estimation of the Path Integral for the Lorenz96 Model

We begin by recalling ideas from the general discussion in Chap. 3 of statistical data assimilation path integrals. The conditional expectation value of a function $G(\mathbf{X})$ on the path \mathbf{X} is given by

$$E[\mathbf{G}(\mathbf{X})|\mathbf{Y}(m)] = \langle G(\mathbf{X}) \rangle = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]} \quad (5.25)$$

The contributions of fluctuations along the path $\mathbf{X} = \{\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_m)\}$ with measurements $\mathbf{Y}(m) = \{\mathbf{y}(t_0), \mathbf{y}(t_1), \dots, \mathbf{y}(t_m)\}$ are accounted for through choices for $G(\mathbf{X})$.

In many, if not most, discussions of statistical data assimilation, a Gaussian assumption about $A_0(\mathbf{X})$ is made at some point. In using the path integral formulation, we are able to directly test this notion. We can evaluate the first through fourth moments of the components of \mathbf{X} , by choosing $G(\mathbf{X})$ to be populated by $[x_a(n)]^\nu$; $a = 1, 2, \dots, D$; $n = 0, 1, \dots, m$; $\nu = 1, 2, 3, 4$. This allows us to estimate the conditional mean trajectory ($\nu = 1$), the RMS variation ($\nu = 2$) about it, as well as the skewness ($\nu = 3$) and kurtosis ($\nu = 4$) of the conditional distribution along the paths.

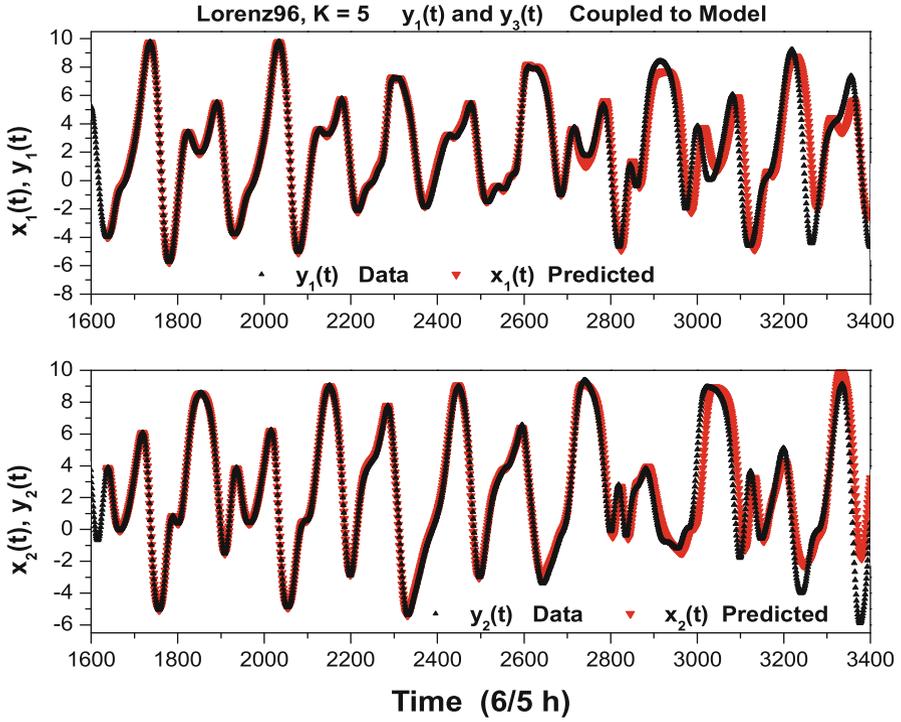


Fig. 5.18 Lorenz96 (Lorenz 1996) model $D = 5$. The “data” for $y_1(t)$ and $y_2(t)$ (up black triangles) along with the prediction for $t > 1600$ from the model for $x_1(t)$ and $x_2(t)$ (red down triangles) are displayed. The initial conditions of all state variables $y_a(t)$ $a = 1, 5$ at $t = 1,600$ were estimated using 1,635 points of $y_1(t)$ and $y_3(t)$ presented to the model. The predictions are accurate for about 1,200–1,400 points beyond the new initial condition at $t = 1,600$, and this is consistent with the estimation of the largest Lyapunov exponent. $x_1(t)$ is an observed state variable; $x_2(t)$ is an unobserved state variable

We can then address the common assumption about the Gaussian nature of the conditional distribution. For a Gaussian distribution the skewness

$$\frac{\langle (x_a(n) - \langle x_a(n) \rangle)^3 \rangle}{\langle (x_a(n) - \langle x_a(n) \rangle)^2 \rangle^{3/2}} \tag{5.26}$$

and kurtosis

$$\frac{\langle (x_a(n) - \langle x_a(n) \rangle)^4 \rangle}{\langle (x_a(n) - \langle x_a(n) \rangle)^2 \rangle^2} - 3 \tag{5.27}$$

both vanish.

Using the model of Lorenz (1996; Lorenz and Emanuel 1998) with $D = 20$ degrees of freedom we have investigated the use of the path integral realized

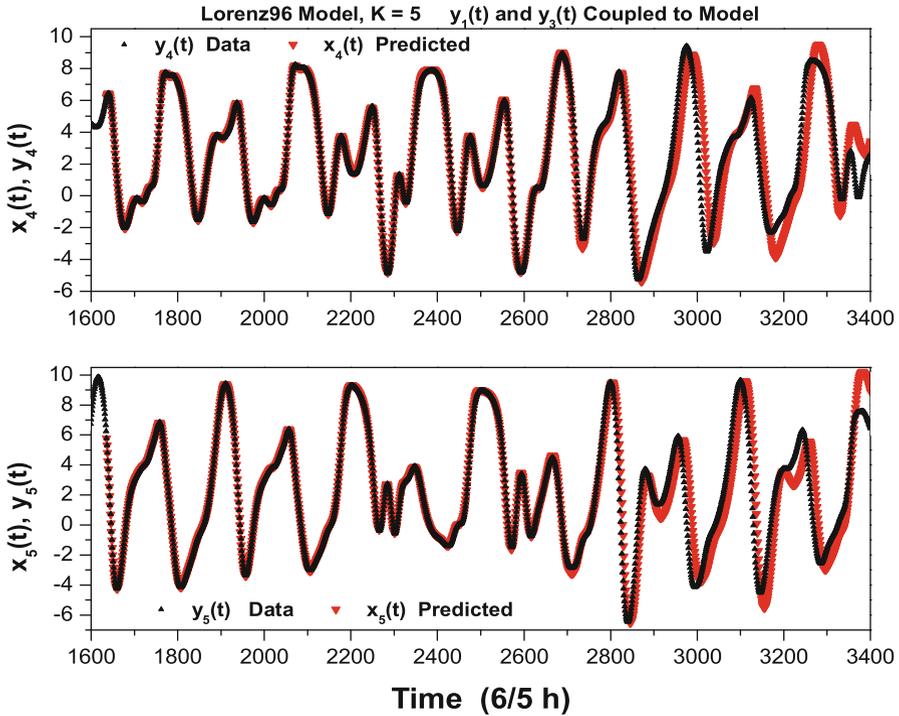


Fig. 5.19 Lorenz96 (Lorenz 1996) model $D = 6$. The “data” for $y_4(t), y_5(t)$ (up black triangles) and the predictions for $t > 1,635$ from the model for $x_4(t), x_5(t)$ (down red triangles) are displayed. The initial conditions at $t = 1,600$ were estimated using 1,635 points of $x_1(t)$ and $x_3(t)$ presented to the model. The predictions are accurate for about 1,200–1,400 points beyond the new initial condition at $t = 1,600$, and this is consistent with the estimation of the largest Lyapunov exponent. Both $x_4(t)$ and $x_5(t)$ are unobserved state variables

numerically using Monte Carlo methods for the estimation of expected values for state variables and parameters, RMS-errors about these expected values, as well as third and fourth moments. The dynamical equations for this model are as given earlier

$$\frac{dx_a(t)}{dt} = x_{a-1}(t)(x_{a+1}(t) - x_{a-2}(t)) - x_a(t) + f = F_a^{(96)}(\mathbf{x}(t), f), \quad (5.28)$$

with $a = 1, 2, \dots, D, x_{-1}(t) = x_{D-1}(t), x_0(t) = x_D(t)$, and $x_{D+1}(t) = x_1(t)$, with $D = 20$ and some choice of $x_a(0)$. We generated a solution $\mathbf{w}(t)$ to the dynamical equations and added Gaussian white noise to the $w_a(t_n)$ to act as our measurements: $\mathbf{y}(n) = \mathbf{w}(n) + \text{noise}$.

We use the approximation to the action discussed in earlier chapters with an implicit Euler representation of the dynamics. We write

$$\begin{aligned}
A_0(\mathbf{X}) &= \frac{R_m}{2} \sum_{n=0}^m \sum_{l=0}^{L_c} (y_l(n) - x_l(n))^2 \\
&\quad + \frac{R_f}{2} \sum_{n=0}^{m-1} \sum_{a=1}^D (x_a(n+1) \\
&\quad - x_a(n) - \frac{\Delta t}{2} [F_a^{(96)}(\mathbf{x}(n), f) + F_a^{(96)}(\mathbf{x}(n+1), f)])^2. \quad (5.29)
\end{aligned}$$

The time step was chosen to be $\Delta t = 0.05$ (Lorenz 1996; Lorenz and Emanuel 1998) corresponding to 6 h in physical time. We selected paths from a space of dimension $(m+1)D + NP = \mathcal{D} + NP$ with $m = 80$ time steps, $D = 20$ state variables, and $NP = 1$ parameter so the path integral has dimension 1,621.

In a Metropolis-Hastings Monte Carlo evaluation of the integral we started from an arbitrary initial path and ignored the first 3×10^5 paths generated, and then used the next 1.2×10^6 paths for the evaluation of the moments of \mathbf{X} . The computation time per path generated is proportional to the number of function evaluations that need to be done per path, proportional to $(8 + NP)(m+1)D$ in this case, and also proportional to the number of random numbers that are generated. We did not try to optimize the Monte Carlo calculations here but used a conventional Metropolis-Hastings with little regard for efficiency or speed of the calculations. We have been concerned with whether the path integral method yields results interesting enough to merit the additional investment in numerical methods to speed up the evaluations of the path integral. In our example calculations with a dimension of $\mathcal{D} + NP = 81 \times 20 + 1 = 1,621$ it takes about 3 ms to generate each path on a contemporary 2 GHz CPU. There is potential for improvement by using more sophisticated algorithms which aim to decrease correlations among paths, but usually with the expense of additional computations (Alexander et al. 2005).

While we report on calculations with $D = 20$, we investigated other model sizes and found little changes in the results except as D grows more observations are needed, as discussed in the next section, and the computation time per path grows linearly with D . For comparison we show below the results of a calculation for $D = 100$.

5.4.1 How Many Observations Are Required?

When model errors are present we have effectively broadened the deterministic delta function of the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ taking the model state $\mathbf{x}(n)$ to $\mathbf{x}(n+1)$. In lowering the resolution from being infinitely accurate in state space, it has reduced the complicated structure of local minima we found in variational principles with no model error. In performing the Monte Carlo evaluation of integrals such as Eq. (5.25), we move around in a \mathcal{D} -dimensional path space

$\mathbf{X} = \{\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_m)\}$ along with an NP -dimensional parameter space seeking to sample the distribution $\exp[-A_0(\mathbf{X})]$. It seems sensible that when this distribution has a well identified set of maxima, associated with minima of $A_0(\mathbf{X})$, the Monte Carlo search with a finite number of iteration steps will best sample paths near the maxima while sampling less well the paths far from the maxima of the distribution.

We know that as R_f becomes very large, we reach the deterministic, no model error, setting. In this limit the search over state and parameter values becomes a constrained search in the space of initial conditions and parameters with dimension $\mathcal{D} + NP$ and is impeded by complex surfaces with many local minima in $A_0(\mathbf{X})$ associated with instabilities on the synchronization manifold (Abarbanel et al. 2008, 2009, 2010) $x_l(n) \approx y_l(n)$; $l = 1, 2, \dots, L$. In our case where R_f is finite, we still see the remnants of this instability in the higher dimensional space of $\mathcal{D} + NP$ dimensions, but it is much less of an issue.

We can turn this idea into a useful tool for estimating the number of observations required to make accurate estimates of model states and parameters by looking at a stochastic process whose limit as it acts multiple times has precisely the distribution function we seek.

Such a stochastic process is given by the Langevin equation (Zinn-Justin 2002)

$$\frac{d\mathbf{X}(s)}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}(s)} + \sqrt{2}\theta(s), \quad (5.30)$$

where $\theta(s)$ is Gaussian white noise with zero mean and variance unity. s is a kind of “time,” really a label for stepping through approximations to the Langevin equation. We touched upon some of the properties of this equation in an earlier chapter. Here we want to examine another aspect of interest.

The Fokker-Planck equation (Lindenberg and West 1990) for the probability distribution $P(\mathbf{X}, s)$ in \mathbf{X} space at “time” s

$$\frac{\partial P(\mathbf{X}, s)}{\partial s} + \frac{\partial}{\partial \mathbf{X}} \left(-\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} P(\mathbf{X}, s) \right) = \frac{\partial^2 P(\mathbf{X}, s)}{\partial \mathbf{X}^2} \quad (5.31)$$

for this process describes how a distribution of paths will develop when each path is evolved according to Eq. (5.30) and tells us that as $s \rightarrow \infty$ the stationary distribution $\exp[-A_0(\mathbf{X})]$ is approached.

This gives an intuitive way to think about how to generate paths to approximate the path integral equation (5.25): starting from any initial path $\mathbf{X}(s = 0)$ the path is evolved in s by moving down the gradient toward local minima [first term in Eq. (5.30)] and fluctuates around those minima [second term in Eq. (5.30)], possibly escaping to other local minima, if the noise is large. In the no noise limit where $\theta \rightarrow 0$ we address the deterministic equation

$$\frac{d\mathbf{X}(s)}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}(s)} \quad (5.32)$$

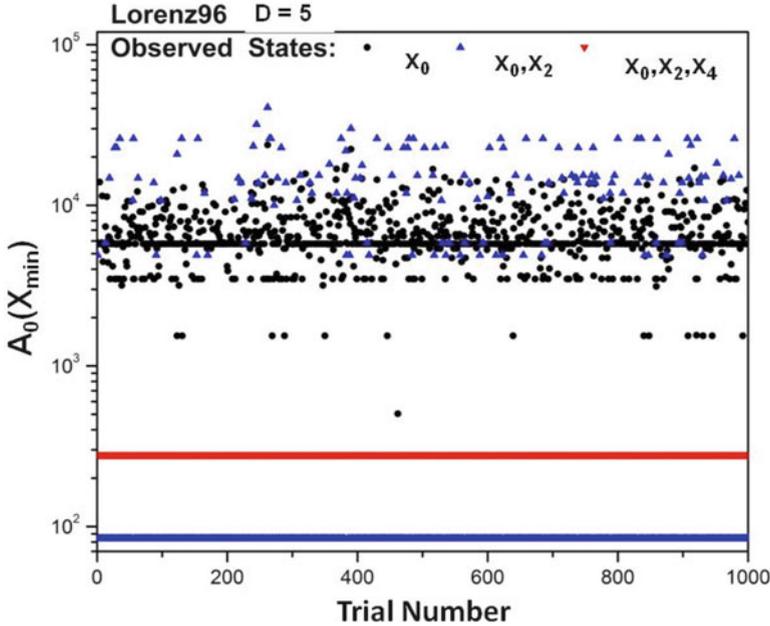


Fig. 5.20 Lorenz96, $D = 5$. $A_0(\mathbf{X}_{\text{final}})$ for 1,000 initial choices of \mathbf{X} allowing the paths to evolve through Eq. (5.32), with the number of observed variables $L = 1, 2$, or 3. For $L = 1$ $A_0(\mathbf{X}_{\text{final}})$ is distributed over a wide range of values. For $L = 2$ there are two clusters of values for $A_0(\mathbf{X}_{\text{final}})$. One is 100 times larger giving little contribution to the path integral. For $L = 3$ there is one cluster of values for $A_0(\mathbf{X}_{\text{final}})$ suggesting the critical value of L is $L_c = 3$

This equation has $A_0(\mathbf{X})$ as a Lyapunov function, since

$$\frac{dA_0(\mathbf{X}(s))}{ds} = - \left| \frac{d\mathbf{X}(s)}{ds} \right|^2, \quad (5.33)$$

As a function of s the solution $\mathbf{X}(s)$ converges to paths $\mathbf{X}(s) \rightarrow \mathbf{X}_{\text{final}}$ at locations in \mathbf{X} space where $\frac{\partial A_0(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{X}_{\text{final}}} = 0$.

If we examine a set of solutions to Eq. (5.32) starting from different $\mathbf{X}(s = 0)$, we will find the collection of minima of $A_0(\mathbf{X})$. Each $\mathbf{X}(s = 0)$ will approach an $\mathbf{X}_{\text{final}}$, and we can develop a picture of the minima in the action by looking at the values of $A_0(\mathbf{X}_{\text{final}})$ for all of the starting paths.

We have followed this idea for the Lorenz96 equations with $D = 5$. In Fig. 5.20 we show the result of 1,000 random selections of $\mathbf{X}(s = 0)$ as a function of the number L of observations. For $L = 1$ we see a plethora of minima of $A_0(\mathbf{X})$ with values ranging from a few times 1,000 to a few times 10,000. This complicated set of possible minima for $A_0(\mathbf{X}_{\text{final}})$ is the reflection of the set of local minima we found in the deterministic, no model error, synchronization error in earlier chapters. When $L = 2$ we see two patterns of minima in $A_0(\mathbf{X})$: one collection of minima a

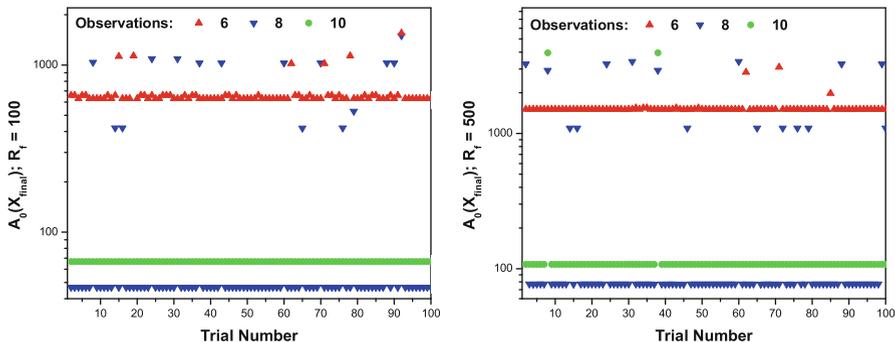


Fig. 5.21 Lorenz96, $D = 20$. $A_0(\mathbf{X}_{\text{final}})$ for 100 initial choices of \mathbf{X} allowing the paths to evolve through Eq. (5.32), with the number of observed variables $L = 6, 8$, or 10 . *Left panel:* $R_f = 100$. *Right panel:* $R_f = 500$

bit less than 100, and another collection of minima with values 50–100 larger. The latter contribute very little to the path integral as it is weighted by $\exp[-A_0(X)]$. This suggests a rather good estimate for the path integral will arise when L is 2 or larger. When we select $L = 3$, there appears only one sharp minimum near $A_0(\mathbf{X}) \approx 300$. This structure reflects the results we encountered in the deterministic variational calculation both in the consideration of CLEs and the examination of the synchronization error as a function of L .

Another example uses the Lorenz96 model again, but now with $D = 20$. We start from 100 different randomly chosen initial paths $\mathbf{X}(s = 0)$ in the $\mathcal{D} + NP$ -dimensional path space. Each state variable component of the initial path was chosen from a uniform distribution over the interval $(-1, 3)$. The component of the path that represents the forcing parameter was chosen from a uniform distribution over the interval $(8, 12)$. In Fig. 5.21 we display the action $A_0(\mathbf{X}_{\text{final}})$ found by integrating Eq. (5.32) for 100 different initial paths in the Lorenz96, $D = 20$ model, for $L = 6, 8$, and 10 observations, $R_m = 50$, with $R_f = 100$ in the left-hand panel and $R_f = 500$ in the right-hand panel. Each initial path is followed using the Langevin equation with no noise until

$$\frac{dA_0(\mathbf{X}(s))}{ds} \approx 0, \tag{5.34}$$

namely, the change in the action with respect to the number of “time” steps is small. This gives us a final path $\mathbf{X}_{\text{final}}$.

For the Metropolis-Hastings Monte Carlo approximation of the path integrals of Eq. (5.25) to be efficient, it is helpful to start with a good approximation of the proposed paths to paths that will be accepted. We will then consider perturbations of these proposed paths to collect good samples of $\exp[-A_0(\mathbf{X})]$ required in doing the high-dimensional path integral Eq. (5.25).

From Figs. 5.20 and 5.21 we see that the surface where $\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}(s)} \approx 0$ can be rather rough (Abarbanel et al. 2008, 2010). It also suggests two ways within the path integral Monte Carlo approach to make the surface of $A_0(\mathbf{X})$ smoother. One way is to increase the number of observed variables. This causes the number of local minima of $A_0(\mathbf{X})$ to be reduced, because information from measurements reduces the number of likely paths. The other way is to increase the uncertainty of the model, by decreasing R_f . The model error is $\propto R_f^{-1/2}$. The figure shows that when $D = 20$, $L = 6$ observations are not enough, but $L = 8$ are enough to smooth the surface. Even though there are still some local minima for $L \geq 8$, the action around these minima is much larger than at the global minimum, so those paths contribute a negligible amount to the path integrals.

5.4.2 Results of Monte Carlo Estimation of the Path Integral for Moments of \mathbf{X} and Model Parameters

In the deterministic case, we know that the minimum number of observations required to remove the instabilities is equal to the number of positive CLEs associated with the synchronization manifold (Abarbanel et al. 2008, 2009, 2010). We have examined this for the Lorenz96 model and found that about $0.4D$ observations (see Table 5.10) are needed to make all CLEs negative (Kostuk 2012). For this reason we selected $L = 8$ for our analysis of the $D = 20$ dimensional Lorenz96 system. In our calculations we chose to “observe” $y_a(n)$ for $a = 0, 3, 5, 8, 10, 13, 15$, and 18 . It is not necessary to have observations at every time step, and in fact here we only provide observations for even n . The missing observation terms are excluded from the conditional mutual information contribution to the action. This is achieved by simply setting $R_m = 0$ at those possible, but unused, measurement times.

We emphasize that this criterion for selecting the required number of observations depends on the model, which is very useful. It also depends on the accuracy of the model, here represented by R_f . The more accurate the model, namely, the larger R_f , the more local minima there will be in $A_0(\mathbf{X})$, because of instabilities on the synchronization manifold of the deterministic problem (Abarbanel et al. 2008, 2009, 2010).

We report here on example calculations using the Lorenz96 model with $D = 20$ and eight observed variables. We used a data assimilation window of 80 steps ($0 \leq t \leq 4$) with observations every other step, followed by a prediction window ($4 < t \leq 6$) with no measurements. $\Delta t = 0.05$ as above. In most cases we considered, the calculations of $\mathbf{x}(n)$ in the prediction window were computed in the deterministic limit ($R_f \rightarrow \infty$) by integrating the model equation forward in time using a fourth-order Runge-Kutta procedure. The integration was done by taking as initial conditions $\mathbf{x}(t_m) = \mathbf{x}(T)$ and parameter values from each sampled path at the end of the assimilation window. This effectively evolves the whole conditional

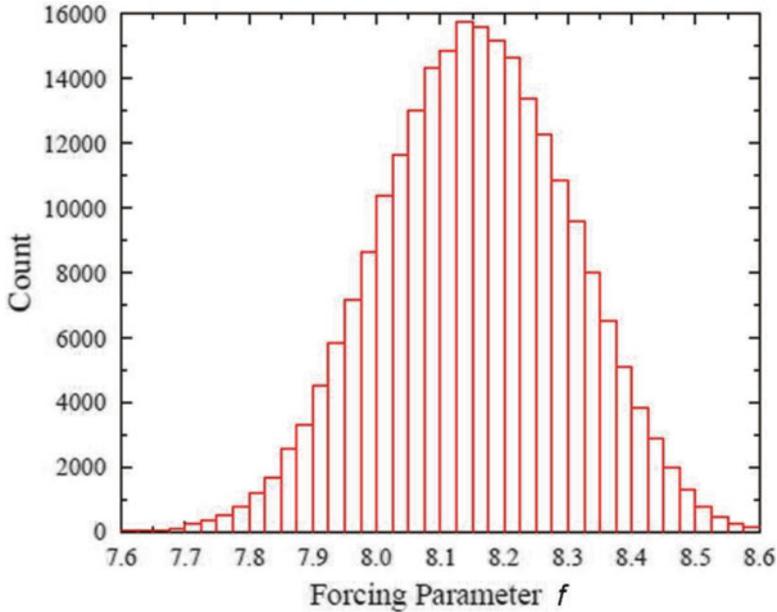


Fig. 5.22 In the Monte Carlo evaluation of the path integral we vary both state variables $x_a(n)$ and parameters. For the Lorenz96 model, the only parameters is the forcing f . For a Lorenz96 model with $D = 20$, we display the marginal distribution of f associated with the Monte Carlo evaluation discussed in the text. The value for the deterministic model was $f = 8.17$

distribution at the end of the assimilation window forward in time into the prediction window. We also did a simpler type of prediction for comparison, where the model equation is integrated forward using only the conditional mean values $\langle \mathbf{x}(t_m) \rangle$ and $\langle \mathbf{p} \rangle$ at the end of the data assimilation window as initial conditions and parameters respectively. These two methods give quite comparable results.

5.4.3 Prediction by Model Equations for $t > t_m$

In performing the Monte Carlo calculation we vary both state variables and parameters seeking paths distributed as $\exp[-A_0(\mathbf{X})]$. In the Lorenz96 model, we have a single parameter. In the calculation for a $D = 20$ Lorenz96 model, we recorded the values of the parameter f for each of the accepted paths, and in Fig. 5.22 we display the marginal distribution of this parameter. The value used in the generation of the data was $f = 8.17$. The marginal distribution of one or more state variables can also be easily extracted from the collection of accepted paths (Figs. 5.23 and 5.24).

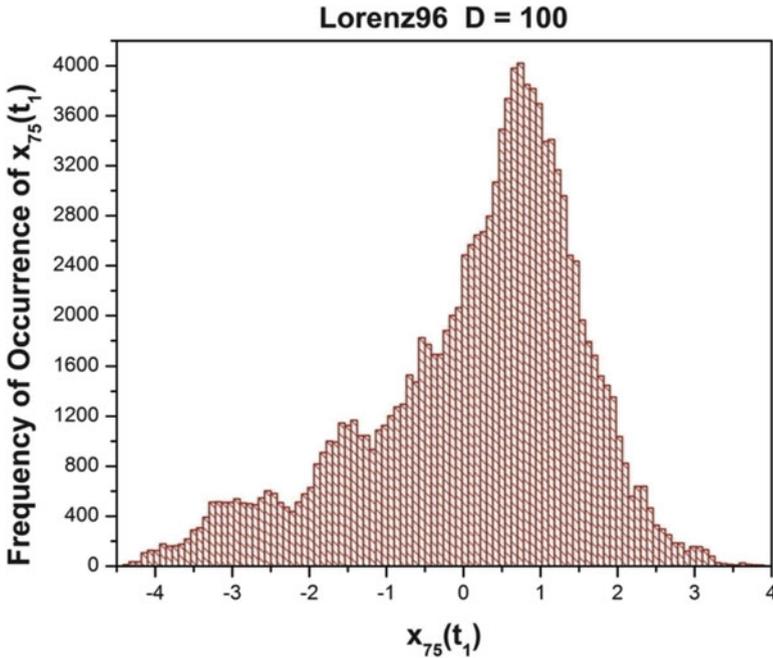


Fig. 5.23 From the metropolis-hastings Monte Carlo procedure, we acquire a collection of accepted paths. These contain the sampled distribution of all state variables at all times in the assimilation window and the distribution of all parameters. Here we display for the $D = 100$ Lorenz96 model the marginal distribution of the state $x_{75}(t_1)$ at the time t_1 . Of interest is the non-Gaussianity of the distribution

Using the values of $\mathbf{x}(T = 4)$ and the parameters f associated with each accepted path, we integrated the equations of motion forward to $t = 6$ using a fourth-order Runge-Kutta procedure. In the **Left Panel** of Fig. 5.25 we display the noisy data for the observed state variable $x_0(t)$, the known clean data value for $x_0(t)$ along with the expected value of $x_0(t)$ and its RMS error about the expected value. These quantities are shown both in the observation or assimilation window $0 \leq t \leq 4$ and in the prediction window $4 < t \leq 6$. In the **Right Panel** we display the skewness and kurtosis of the distributions in both the observation and the prediction regions. For the unobserved state variable $x_{19}(t)$, we display the same quantities in Fig. 5.26, except there are no observations to show.

The two state variables we have displayed behave in the same manner as the other 18 variables, observed and unobserved. We generated the “data” (solid black lines) by integrating the model equations with $f = 8.17$ and some choice of initial conditions. This gives us all the state variables of the observed system. We then generated “observations” (blue circles) from eight of the state variables by adding Gaussian noise with standard deviation $\sigma_m = 0.353$ to the true path. There is no correlation in the noise at different time steps or among different variables. We selected $R_m = 1/\sigma_m^2 \approx 8$ and $R_f = 100$ for these calculations.

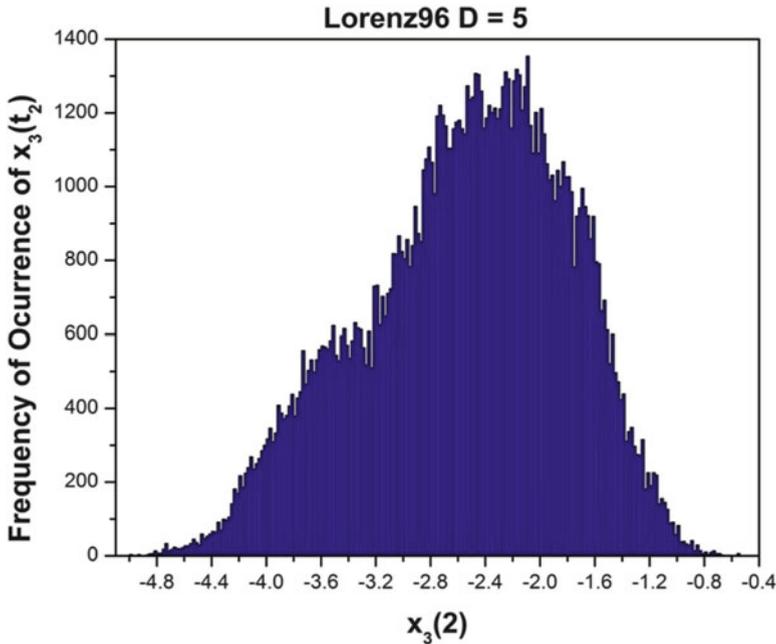


Fig. 5.24 From the metropolis-hastings Monte Carlo procedure, we acquire a collection of accepted paths. These contain the sampled distribution of all state variables at all times in the assimilation window and the distribution of all parameters. Here we display for the $D = 5$ Lorenz96 model the marginal distribution of the state $x_3(2)$ at the time t_2 . Of interest is the non-Gaussianity of the distribution

The expected state estimates track the known path quite well in both the assimilation and the prediction windows for the observed as well as the unobserved state variables. The uncertainties for the predicted states grow in time, because the largest Lyapunov exponent of the model is positive, about 0.9 in units of inverse time (Kostuk 2012). This means that at the end of the prediction period, $t = 6$, the uncertainties should be about six times as large as the uncertainties at the end of the assimilation window, $t = 4$.

The accuracy with which the expected state variables track the known data indicates that both the eight observed states as well as the twelve unobserved states have been accurately estimated at the end of the observation window. From the displayed examples $x_0(t)$ and $x_{19}(t)$ we can also see that the estimations within the observation window are also quite accurate.

In the example of Figs. 5.25 and 5.26, the skewness and kurtosis are both close to zero in the assimilation window. They are slightly smaller in magnitude for $R_f = 100$ than for $R_f = 500$ (not shown). This suggests that the conditional distributions are nearly Gaussian during the assimilation window, probably because of the influence of the measurements. The ratio $\frac{R_m}{R_f}$ is the determining factor.

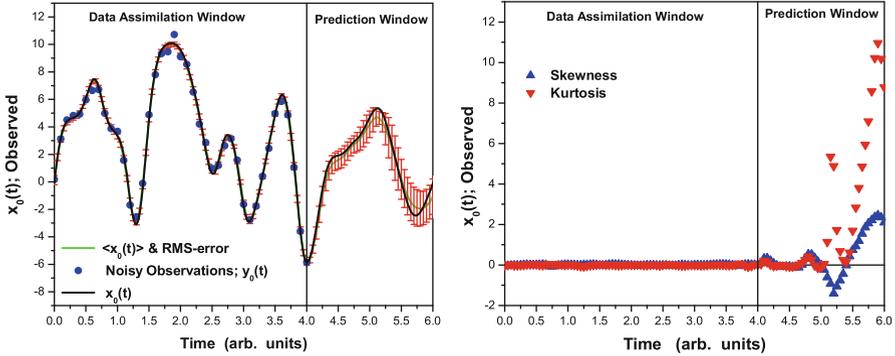


Fig. 5.25 Details of the estimation using the Monte Carlo evaluation of the data assimilation path integral for the Lorenz96 model with $D = 20$. We display the estimation and prediction of the observed variable $x_0(t)$. *Left panel*: conditional mean (green line) and RMS errors (red) along with Gaussian distributed noisy observations (blue solid circles) and the known $x_0(t)$ (solid black line). *Right panel*: skewness (blue up triangles) and kurtosis (red down triangles) of $x_0(t)$. These are Monte Carlo estimates from the path integral for the Lorenz96 model with $D = 20$; $R_m = 8$, $R_f = 100$. The assimilation window is $0 \leq t \leq 4$, and the prediction window is $4 < t \leq 6$. There are no observations in the prediction window. The parameter estimate is $f = 8.25 \pm 0.09$. Predictions for $t > 4$ are made with a fourth-order Runge–Kutta procedure using information on the estimated parameter and state variables at $t = 4$

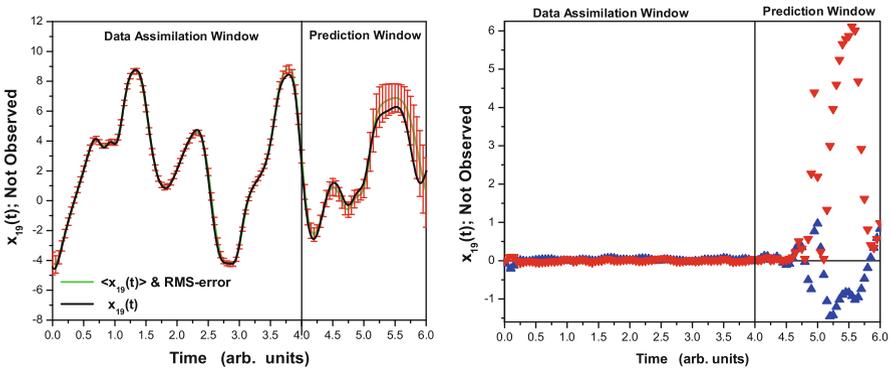


Fig. 5.26 Details of the estimation using the Monte Carlo evaluation of the data assimilation path integral for the Lorenz96 model with $D = 20$. We display the estimation and prediction of the unobserved variable $x_{19}(t)$. *Left panel*: conditional mean (green line) and RMS errors (red) along with known $x_{19}(t)$ (solid black line). No observations here. *Right panel*: skewness (blue up triangles) and kurtosis (red up triangles) of $x_{19}(t)$. These are Monte Carlo estimates from the path integral for the Lorenz96 model with $D = 20$; $R_m = 8$, $R_f = 100$. The assimilation window is $0 \leq t \leq 4$, and the prediction window is $4 < t \leq 6$. The parameter estimate is $f = 8.25 \pm 0.09$. Predictions for $t > 4$ are made with a fourth-order Runge–Kutta procedure using information on the estimated parameter and state variables at $t = 4$

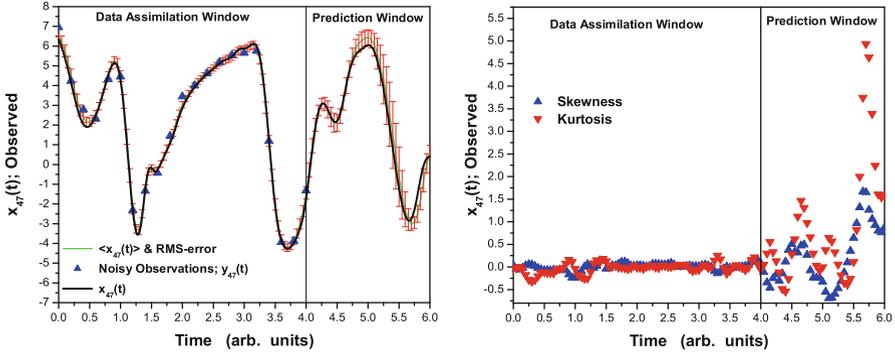


Fig. 5.27 Details of the estimation using the Monte Carlo evaluation of the data assimilation path integral for the Lorenz96 model with $D = 100$. We display the estimation and prediction of the observed variable $x_{47}(t)$. *Left panel*: conditional mean (*dashed line*) and RMS error along with Gaussian distributed noisy observations (*open circles*) and known $x_{47}(t)$ (*solid line*). *Right panel*: skewness (*open triangles*) and kurtosis (*solid triangles*) of $x_{47}(t)$. In the action we selected $R_m = 8$, $R_f = 100$. The assimilation window is $0 \leq t \leq 4$, and the prediction window is $4 < t \leq 6$. Predictions for $t > 4$ are made with a fourth-order Runge–Kutta procedure using information on the estimated parameter and state variables at $t = T = 4$. The parameter used in generating the clean data was $f = 8.17$, while the estimate from the Monte Carlo calculation using noisy data is $f = 8.32 \pm 0.044$

When this is sizeable, the Gaussian errors in the observations are important. When this ratio goes to zero, in the deterministic or zero model error limit, the non-Gaussian part of the action is dominant. When the assimilation window ends, the distribution is evolved according to the nonlinear dynamics of the model, and so it becomes much more non-Gaussian and less localized because of chaos. The distribution can become quite complicated with the regions containing the most probable paths no longer contiguous in path space.

This suggests that as the models of the observed process become better and better, namely, as model error is reduced, the role of the nonlinear, non-Gaussian elements of the action $A_0(\mathbf{X}, \mathbf{Y}(m))$ will be more and more important. Approximations to the assimilation of information from measurements based on Gaussian assumptions may become less valuable in this circumstance.

We performed calculations such as those reported here for a variety of values of the dimension D of the model. The results were much the same as for $D = 20$ and for comparison we show for $D = 100$ the conditional expectation value of an observed state $x_{47}(t)$. This is in Fig. 5.27. We used 40 observations at each time point where observations were made. The 60 unobserved states and the parameter f were estimated using the Monte Carlo approximation to the path integrals. The set of estimated values of the observed and unobserved states at $T = 4$ associated with each accepted path along with the parameters associated with that path were used as initial conditions and parameter values for prediction into $t > 4$ using a fourth-order Runge-Kutta algorithm. The predicted paths were averaged to yield an expected prediction, and the RMS error about this mean was evaluated. These results are displayed in Fig. 5.27.

5.4.4 Non-Gaussian Measurement Error

In the use of the path integral method, or any other data assimilation approach actually, we do not know the statistics of the error in the measurements, and while the assumption that they are Gaussian is common, it is by no means necessary. To examine the implication of selecting another distribution of the errors, we represented the measurement errors by a generalized Lorentzian distribution (also known as a Cauchy distribution)

$$P(z) \propto \frac{1}{(1+z^2)^4}. \quad (5.35)$$

This replaces the conditional mutual information term in the action by

$$-\sum_{n=0}^m CMI(\mathbf{x}(n), \mathbf{y}(n) | \mathbf{Y}(n-1)) = 4 \sum_{n=0}^m \sum_{l=1}^L \log \left(1 + \frac{R_m}{2} (y_l(n) - x_l(n))^2 \right). \quad (5.36)$$

We chose the generalized Lorentzian distribution as it, too, is an approximation to a delta function that would here represent zero noise in the measurements. Selecting the power 4 here was only to assure that many moments of the distribution were finite while yielding a power law tail rather than the exponential tail of a Gaussian. We tried other powers such as 6, and it appeared to make no difference.

Our goal was to select a wrong distribution for the noisy measurements and determine if it mattered very much. We provided measurements with Gaussian noise. The effect is a somewhat larger, though still small, skewness and kurtosis in the assimilation window.

Figure 5.28 shows the results of this calculation, using the same measurement data as the other two examples, and with $R_f = 100$, $R_m = 8$. This change does not make a substantial difference for the values of R_m and R_f we used, except that the conditional distributions of the observed variables in the assimilation window are slightly less consistent with a Gaussian than before. The exploration of the effect of non-Gaussian measurement noise distributions can be explored in a straightforward manner through the use of path integral Monte Carlo methods.

5.4.5 How Often Should One Make Measurements?

Additionally, we addressed the question of how often observations are required in the data assimilation window. The answer to this is likely to be model dependent, but for the Lorenz96 model with $D = 20$, we addressed this question by doing the Metropolis-Hastings Monte Carlo calculation for several different frequencies of making observations in the assimilation window, ranging from observations at all 80 time points down to observations at only three time points. In all cases eight

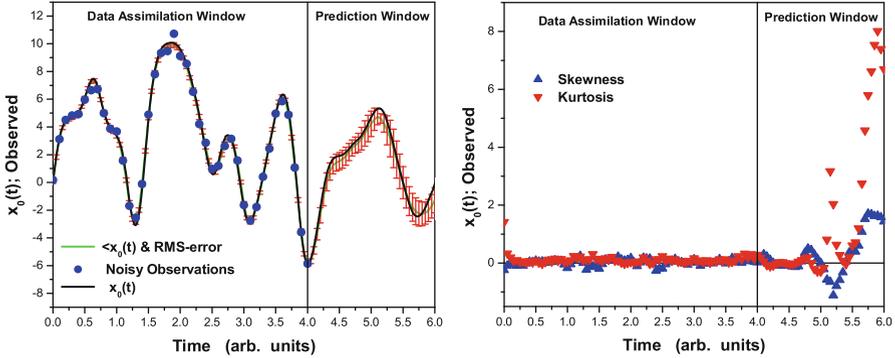


Fig. 5.28 Details of the estimation and prediction using the Monte Carlo evaluation of the data assimilation path integral for the Lorenz96 model with $D = 20$ when we take the error in the measurements to be distributed as a Lorentzian; $P(z) \propto (1 + z^2)^{-4}$. The noise added to the clean data in this twin experiment was Gaussian. We display the estimation and prediction of the observed variable $x_0(t)$. *Left panel:* conditional mean (*dashed line*) and RMS error along with Gaussian distributed noisy observations (*open circles*) and known $x_0(t)$ (*solid line*). *Right panel:* skewness (*open triangles*) and kurtosis (*solid triangles*) of $x_0(t)$. These are Monte Carlo estimates from the path integral for the Lorenz96 model with $D = 20$; $R_m = 8$, $R_f = 100$. The assimilation window is $0 \leq t \leq 4$, and the prediction window is $4 < t \leq 6$. In these calculations the conditional mutual information term in the action represents the measurement errors as the generalized Lorentzian, Eq. (5.36). The actual distribution of errors is Gaussian. The parameter estimate is $f = 8.24 \pm 0.08$. Predictions for $t > 4$ are made with a fourth-order Runge–Kutta procedure using information on the estimated parameter and state variables at $t = 4$

variables were observed; it is only the data sampling rate that is being changed. As a diagnostic we compare the conditional mean $\langle x_a(n) \rangle$ with the true trajectory $w_a(n)$ by computing the RMS error and dividing by the RMS variation over time of $\langle x_a(n) \rangle$. We define the diagnostic

$$\frac{\sqrt{\frac{1}{m} \sum_{n=1}^m (w_a(n) - \langle x_a(n) \rangle)^2}}{\sqrt{\frac{1}{m} \sum_{n=1}^m (\langle x_a(n) \rangle - \bar{x}_a)^2}}, \tag{5.37}$$

where $\bar{x}_a = \frac{1}{m} \sum_{n=1}^m \langle x_a(n) \rangle$. We plot this for one observed variable, $x_8(t)$, and one unobserved variable, $x_9(t)$, in Fig. 5.29. The accuracy of the estimations remains good for a wide range of number of measurements, then degrades rapidly as the number becomes small. We have not made a systematic study of this effect, but we suspect that as the time between measurements becomes larger than the inverse of the largest Lyapunov exponent of the model, the quality of these estimations will degrade.

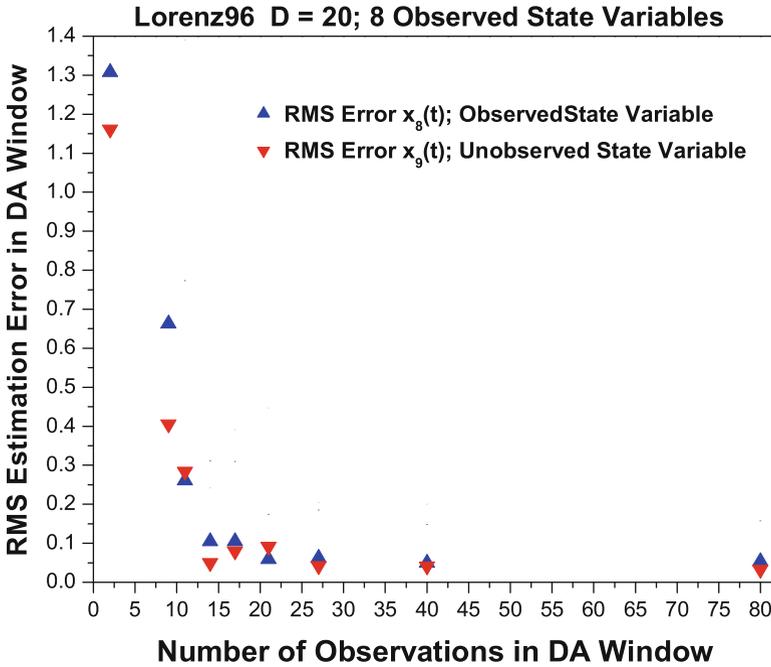


Fig. 5.29 RMS error of estimated state variables over the data assimilation window. The maximum on the x -axis, 80, means that there were eight observations at 80 different equally spaced times in the data assimilation window $0 \leq t \leq 4$. We report the diagnostic defined in Eq. (5.37) for one observed variable, $a = 8$, and one unobserved variable $a = 9$. The *up blue triangle* shows the error for the observed state variable $x_8(t)$. The *down red triangle* shows the error for the unobserved state variable $x_9(t)$

5.5 Shallow Water Equations

In this section we turn to the analysis of statistical data assimilation in a geophysical flow, a one-layer shallow water equation described in detail below. Three fields define the flow: horizontal velocity in the x and y directions and the height of the fluid over a fixed base. We create an $N \times N$ grid and discretize the flow to $D = 3N^2$ ordinary differential equation. We carry out a twin experiment by first generating data in a chaotic flow regime and then asking how many and which observations among the D we must observe to permit synchronization of the data with the output of the model generated with different initial conditions then used for generating the data. Once this critical number of observations L_c is established and the identification of the particular observations comprising the required L_c , we perform a statistical data assimilation using the path integral evaluated via the Monte Carlo method already described. As all parameters are set, we seek to accurately estimate the unobserved state variables in this. The accuracy of the state variable estimation is tested, validated or invalidated, by predicting beyond the observation

window, and we find that when the number of observations and the appropriate observations are used, so that the data and the model output synchronize, the predictions are very accurate. In other cases, when synchronization is absent, the predictions are not accurate.

We display an example where the accuracy of estimation in the observation window appears excellent though synchronization does not occur for that selection of observations, and prediction is not good for that situation. We also show an example where estimation in the observation window is inaccurate and prediction is equally inaccurate. If one is using statistical data assimilation to establish a sense of the state of the model at the termination of observations, that alone will not be informative if the observations are not among the collections that exhibit synchronization.

We first describe the physics of the fluid model, a shallow layer of fluid in a rectangular section of a rotating earth, where the standard β -plane approximation is made to the local rotation around the vertical direction (Pedlosky 1987; Vallis 2006; Gill 1982).

5.5.1 One-Layer Shallow Water Flow

In atmospheric and ocean models of flows related to weather and climate, there is a natural simplifying approximation arising from the small scale of vertical dimensions compared to horizontal spatial dimensions on the earth's surface. The vertical scale L_V is order of 10 km in the atmosphere, and a bit smaller in the ocean, while the horizontal scale L_H is 100's to 1000's of kilometers. This results in large horizontal velocities $u(x, y, z, t)\hat{x} + v(x, y, z, t)\hat{y}$ relative to the vertical velocity $w(x, y, z, t)\hat{z}$ in the shallow layer of fluid: $w \approx \frac{L_V}{L_H}u$. This is illustrated in Fig. 5.30.

We proceed along with the detailed discussion presented by Pedlosky (1987; Gill 1982) of the dynamics of a shallow waterflow in a Cartesian $(x, y, z) = (\mathbf{r}, z)$ frame rotating about the z -axis at a rate $\frac{f(\mathbf{r})}{2}$. This is to be thought of as a patch, perhaps a few 100 km on a side at mid-latitudes on the earth's surface. The dynamics of the flow are determined by the horizontal velocity $u(\mathbf{r}, z, t)\hat{x} + v(\mathbf{r}, z, t)\hat{y}$ and the time varying surface height $h(\mathbf{r}, t)$. The density of the fluid is taken as a constant ρ and gravity force is $-\rho g\hat{z}$. The momentum dynamics in the vertical direction is dominated by hydrostatic balance

$$0 = -\frac{\partial p(\mathbf{r}, z, t)}{\partial z} - \rho g, \quad (5.38)$$

where the vertical gradient of the pressure is balanced by the force of gravity. If one has significant variation in density, this is usually represented by many shallow layers in each of which the fluid density is taken constant.

The vertical momentum equations neglect accelerations associated with vertical motion, and we have for the pressure

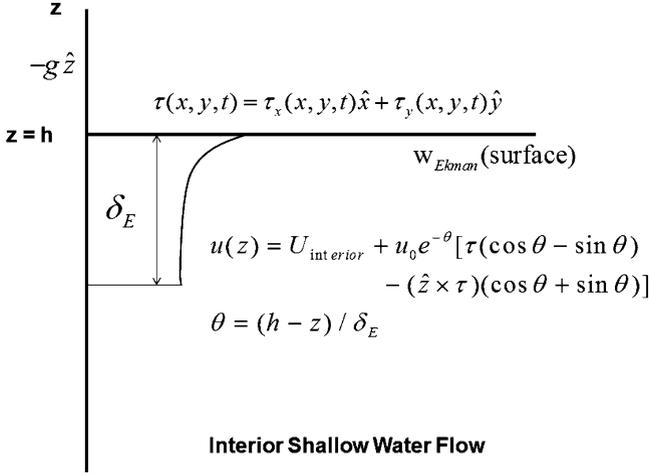


Fig. 5.30 Illustration of the Ekman pumping of the interior, shallow waterflow which transfers forcing from the wind stress $\boldsymbol{\tau}(\mathbf{r})$ at the surface. A boundary layer of thickness $\delta_E = \sqrt{\frac{2A}{f}}$ [see Eq. (5.44)] in which the horizontal velocity \mathbf{u} varies rapidly in z is formed. \mathbf{u} decays exponentially from its value at the surface $z = h(\mathbf{r}, t)$ and rotates its direction. z and h are measured in units of the scale L_V of vertical depths in the fluid

$$p(\mathbf{r}, z, t) = \rho g(h(\mathbf{r}, t) - z), \quad (5.39)$$

taking the pressure at the surface $z = h(\mathbf{r}, t)$ equal zero.

We account for frictional forces in two different ways:

- Add Newtonian shear friction and bottom ($z = 0$) friction to the horizontal momentum equations.
- Represent the forcing of the fluid as wind stresses $\boldsymbol{\tau}(\mathbf{r}, t) = \tau_x(\mathbf{r}, t)\hat{x} + \tau_y(\mathbf{r}, t)\hat{y}$ at the fluid surface $z = h(\mathbf{r}, t)$. To meet the boundary condition at the surface where pressure and stresses must be continuous (Batchelor 1967), a thin layer near the surface is identified in which the flow in the interior transitions to that dictated by the boundary conditions. This kind of boundary layer is familiar in dissipative fluid flows (Schlichting and Gersten 2000).

In the interior flow we have the momentum equations (Batchelor 1967) (Navier-Stokes equations) for $\mathbf{u}(\mathbf{r}, t) = u(\mathbf{r}, t)\hat{x} + v(\mathbf{r}, t)\hat{y}$ with $\mathbf{r} = (x, y)$

$$\begin{aligned} & \frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial t} + \mathbf{u}(\mathbf{r}, t) \cdot \nabla \mathbf{u}(\mathbf{r}, t) \\ &= -\nabla p(\mathbf{r}, z, t) + \mathbf{u}(\mathbf{r}, t) \times \hat{z} f(\mathbf{r}) + \text{friction} + \text{body forcing} \\ &= -g \nabla h(\mathbf{r}, t) + \mathbf{u}(\mathbf{r}, t) \times \hat{z} f(\mathbf{r}) + \text{friction} + \text{body forcing}. \end{aligned} \quad (5.40)$$

Mass conservation, when ρ is a constant, is

$$\frac{\partial w(\mathbf{r}, z, t)}{\partial z} + \nabla \cdot \mathbf{u}(\mathbf{r}, t) = 0, \quad (5.41)$$

which can be integrated in the z -direction to yield

$$w(\mathbf{r}, z, t) = -z \nabla \cdot \mathbf{u}(\mathbf{r}, t) + W(\mathbf{r}, t), \quad (5.42)$$

and $W(\mathbf{r}, t)$ will be determined by the vertical flow at the bottom of the thin layer between the interior flow and the surface.

To represent the stress from wind above the fluid surface the stress tensor is parametrized by an effective viscosity A

$$\tau_x = \rho A \frac{\partial u}{\partial z} \quad \tau_y = \rho A \frac{\partial v}{\partial z}. \quad (5.43)$$

In the thin layer where friction transports energy from the wind stress to the interior flow, the horizontal velocity varies rapidly with z . Using these forms for the stress, the horizontal velocities are found to vary in the vertical on a scale called the Ekman thickness

$$\delta_E = \sqrt{\frac{2A}{f}}, \quad (5.44)$$

which is approximately 10's of meters in ocean flows.

Integrating the flow through the thin layer results in a horizontal flux of mass that requires a vertical flow at the surface. Combining this with the definition of $w(\mathbf{r}, h(\mathbf{r}, t)) = \frac{dh(\mathbf{r}, t)}{dt}$ when there is no forcing yields the dynamical equation for $h(\mathbf{x}, t)$

$$\frac{\partial h(\mathbf{r}, t)}{\partial t} + \nabla \cdot (\mathbf{u}(\mathbf{r}, t)h(\mathbf{r}, t)) = -\hat{z} \cdot \text{curl} \left\{ \frac{\boldsymbol{\tau}(\mathbf{r}, t)}{\rho f(\mathbf{r})} \right\}. \quad (5.45)$$

Summarizing these equations and making the bottom and interior friction explicit, we have

$$\begin{aligned} \frac{\partial h(\mathbf{r}, t)}{\partial t} + \nabla \cdot (h(\mathbf{r}, t)\mathbf{u}(\mathbf{r}, t)) &= -\hat{z} \cdot \text{curl} \left(\frac{\boldsymbol{\tau}(\mathbf{r}, t)}{\rho f(\mathbf{r})} \right) \\ \frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial t} + \mathbf{u}(\mathbf{r}, t) \cdot \nabla \mathbf{u}(\mathbf{r}, t) &= -g \nabla(h(\mathbf{r}, t)) + A \nabla^2 \mathbf{u}(\mathbf{r}, t) - \boldsymbol{\epsilon} \mathbf{u}(\mathbf{r}, t) \\ &\quad + \mathbf{u}(\mathbf{r}, t) \times \hat{z} f(\mathbf{r}) + \text{body forcing}. \end{aligned} \quad (5.46)$$

This is a compact set of fluid flow equations with forcing both in the interior flow and from wind stresses transferred through a thin upper boundary layer called the Ekman layer. The flow through the Ekman layer pumps mass and momentum into the interior layer adding forcing in the $h(\mathbf{r}, t)$ equations to any intrinsic body forces that could arise from other sources.

If the rotation rate is large enough so that the Rossby number (Pedlosky 1987; Vallis 2006; Gill 1982)

$$Ro = \frac{U}{fL_H}, \quad (5.47)$$

where U is the scale of the horizontal velocity, is small, then the dominant feature of the interior flow is a balance between gravity forcing $-g\nabla h(\mathbf{r}, t)$ and the Coriolis force $\mathbf{u}(\mathbf{r}, t) \times \hat{\mathbf{z}}f(\mathbf{r})$ leading to geostrophic flow

$$\mathbf{u}_{\text{geostrophic}}(\mathbf{r}, t) = \frac{g\hat{\mathbf{z}} \times \nabla h(\mathbf{r}, t)}{f(\mathbf{r})}. \quad (5.48)$$

Corrections to this can be expressed in an expansion in Ro which is about 0.1 in the atmosphere and 0.01 in the ocean (Vallis 2006; Gill 1982).

Interestingly in one of those serendipitous connections these equations also describe the flow of a beam of charged particles in strong crossed electric and magnetic fields: the electric force is analogous to the gravity driving and the magnetic force is analogous to the Coriolis force. Geostrophic flow is called “ $E \times B$ ” drift in the charged particle motion (Northrop 1963).

5.5.2 Statistical Data Analysis for the Shallow Water Equations

Our interest in these equations arises with the question: how many measurements on the three fields $\{\mathbf{u}(\mathbf{r}, t), v(\mathbf{r}, t), h(\mathbf{r}, t)\}$ at each time t_n in an observation interval $\{t_0, t_1, \dots, t_m = T\}$ are required to allow accurate estimation of the parameters in the model and the full state of the model at $t = T$ and then allow for accurate prediction for $t > T$?

We analyze the situation where there are no body forces in the interior, so the fluid is forced solely by Ekman pumping from the surface wind stresses through the boundary layer. We selected mid-latitude dynamics on a β plane with coordinates $\mathbf{r} = (x, y)$ and rotating at $\mathbf{f}(\mathbf{r})/2$ about the z -axis, which means $f(\mathbf{r}) = f_0 + \beta y$. We also chose as the forcing at the surface of the fluid a stress representing wind in the East–West direction varying with the North–South coordinate: $\tau_x(\mathbf{r}) = F_W \cos(2\pi y)$. Finally we write $h(\mathbf{r}, t) = H_0 + \eta(\mathbf{r}, t)$. The parameters for the flow are reported in Table 5.11. Most parameters are taken from the literature on shallow water flows; however, the forcing strength F_W was selected so the solutions to the

Table 5.11 Parameters used in the generation of the shallow water “data” for the twin experiment

Parameter	Physical quantity	Value in twin experiments
Δt	Time step	36 s
ΔX	East–west grid spacing	50 km
ΔY	North–south grid spacing	50 km
H_0	Equilibrium depth	5.1 km
f_0	Rotation rate	$5 \times 10^{-5} \text{ s}^{-1}$
β	Coriolis parameter	$2.0 \times 10^{-11} \text{ (s-m)}^{-1}$
F_W	Wind stress	$0.2 \text{ m}^2/\text{s}^3$
ν	Effective viscosity	$10^{-4} \text{ m}^2/\text{s}$
ϵ	Rayleigh friction	$2 \times 10^{-8} \text{ s}^{-1}$

shallow water equations on a 16×16 grid were chaotic. We return below to how we know the flow is chaotic.

If we place the flow on an $N \times N$ grid for $\mathbf{r}=(x, y)$, we have $D=3N^2$ degrees of freedom expressed in ordinary differential equations for $\{u(i, j, t), v(i, j, t), h(i, j, t)\}$ on the grid labeled by locations $(x, y) = \{i \Delta x, j \Delta y\}$; $i, j = 1, 2, \dots, N\}$ and $(\Delta x, \Delta y)$ the spatial step size on the grid.

To determine the number of required measurements for a selected flow regime, we want to know how many observations L_c at each observation time are needed to produce CLEs that are all negative, indicating the stability of the synchronization manifold. We use L_c to indicate the critical number of observations leading to all negative CLEs on the synchronization manifold.

We have identified three ways to determine L_c :

- Couple $L = 1, 2, \dots, L_c$ data time series into the shallow water flow model through $u_{al}(y_l(n) - x_l(n))$. Then use the model data sets construct the $D \times D$ Oseledec matrix for $L = 1, 2, \dots, L_c$. Determine the largest CLE for the Oseledec matrix for $L = 1, 2, \dots, L_c$ and establish when the largest CLE becomes negative.
- Introduce $L = 1, 2, \dots, L_c$ data time series into the action for the shallow water flow model and using the model data sets construct the action $A_0(\mathbf{X}, \mathbf{Y}(m))$. For each L solve the Langevin equation

$$\frac{dA_0(\mathbf{X}(s))}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial \mathbf{X}(s)} \tag{5.49}$$

for many $\mathbf{X}(s = 0)$ going from this initial choice to $\mathbf{X}_{\text{final}}$. By examining $A_0(\mathbf{X}_{\text{final}})$ determine when one or a few sharp minima appear.

- Couple $L = 1, 2, \dots, L_c$ data time series, as in the first method, into the shallow water flow model and determine when the synchronization error between the data time series and the equivalent model output goes to zero as a function of L . One must establish not only **how many** data time series L_c are necessary, but we must also identify **which** data time series are required.

Of these three methods the third is the most computationally efficient as $D = 3N^2$ becomes large. The first requires accurate handling of $3 \times N^2$ by $3 \times N^2$ matrices which can be quite ill-conditioned. The second requires solving a differential equation in $3 \times N^2 \times (m + 1)$ dimensions many times to exhibit the distribution of minima of $A_0(\mathbf{X})$ for each $L = \{1, 2, \dots, L_c\}$. The third requires repeated solutions of the dynamical equations for different values of L and the coupling strength. This is the approach we pursue here.

Once one has established L_c and determined which L_c measurements at grid points are needed, then one may check the utility and physical viability of the answer by using those L_c measurements to determine the full state $\mathbf{x}(T)$ at $t = T$ and then predicting forward in time from $\mathbf{x}(T)$ as initial conditions. If $L < L_c$, predictions are likely to be inaccurate as the estimation of the full model state at T is probably not very good. For $L \geq L_c$, predictions could be accurate within the intrinsic limitations of a chaotic flow, if the model output is chaotic.

5.5.3 Generating the Data

We discretized the shallow water equations on a regular grid following (Sadourny 1975). We selected an enstrophy (Pedlosky 1987; Gill 1982) conserving shallow water system on a staggered C grid. On this grid we locate the dynamical variables as follows (see Fig. 5.31):

- Pressure (per unit mass) $P(\mathbf{r}, t)$ and height $h(\mathbf{r}, t)$ information is on gridpoints.
 $P(\mathbf{r}, t) = gh(\mathbf{r}, t)$.
- $u(\mathbf{r}, t)$ is defined on lines connecting adjacent horizontal (East–West) gridpoints.
- $v(\mathbf{r}, t)$ is defined on lines connecting adjacent vertical (North–South) gridpoints.

Using the parameters in Table 5.11 we generated data on a 16×16 grid using the Sadourny (1975) enstrophy conserving discretization scheme. Periodic boundary conditions were imposed on the flow. After integrating these equations for several thousand model time hours and discarding the transient flow, we display in Fig. 5.32 a snapshot of the flow at the end of the assimilation window, $t = 40$ h.

The flow generated with these parameters is chaotic. We demonstrate this by evaluating the largest Lyapunov exponent for the $3 \times N^2 = 768$ dimensional system. To accomplish this we selected two sets of very close initial conditions, $\mathbf{x}_{\text{set } 1}(0)$ and $\mathbf{x}_{\text{set } 2}(0)$, and then we integrated the equations forward in blocks of 4,000 time steps of 0.01 h each. During each block of 40 h, we tracked the RMS distance between the two slightly different solutions to the shallow water equations and evaluated the largest Lyapunov exponents by approximating the growth of this error as $|\mathbf{x}_{\text{set } 1}(t) - \mathbf{x}_{\text{set } 2}(t)| \approx |\mathbf{x}_{\text{set } 1}(0) - \mathbf{x}_{\text{set } 2}(0)| \exp[\text{LE}(t/1 \text{ h})]$. We then recorded LE for each block of time steps.

Eventually this difference, selected to be small at $t = 0$, grows quite large, so we follow standard practice (Kantz and Schreiber 2004) and rescale the value at a later time to be quite small again. Then we continue our integration once again. Since

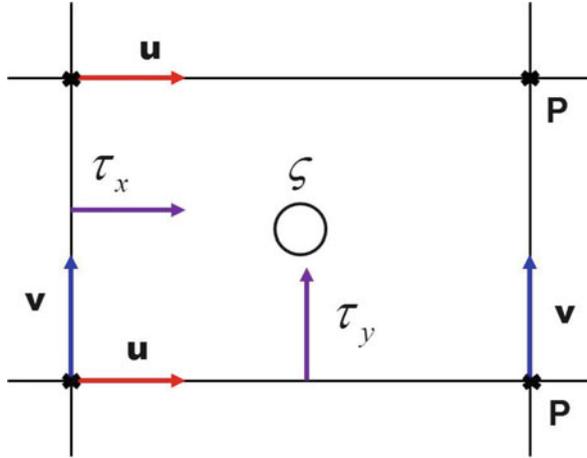


Fig. 5.31 Display of the location on the $N \times N$ grid ($N = 16$ in our calculations) of the dynamical variables for the shallow water flow forced by Ekman pumping. The pressure (per unit mass) $P(\mathbf{r}, t)$ and height $h(\mathbf{r}, t)$ information are located on gridpoints. $P(\mathbf{r}, t) = gh(\mathbf{r}, t)$. The meridional velocity $u(\mathbf{r}, t)$ is defined on lines connecting adjacent horizontal (East–West) gridpoints, and the zonal velocity $v(\mathbf{r}, t)$ is defined on lines connecting adjacent vertical (North–South) gridpoints. $\zeta(\mathbf{r}, t) = (\text{curl } \mathbf{u}(\mathbf{r}, t))_3$

the Lyapunov exponents are defined by a linearized flow on the tangent plane of the dynamics, the rescaling causes no harm and enables numerical stability.

After rescaling we tracked the two orbits for another block of 40 h in time, and we continued in this manner for 80 blocks of time each of 40 h of duration. In Fig. 5.33 we present the estimated largest Lyapunov exponents LE from each of the horizontal velocity and height fields. These estimates vary a small amount from each other as the time window over which nearby orbits is tracked is only 40 h and also because each rescaling starts an orbit from different locations on the system attractor, and the local Lyapunov exponents, for finite time segments, are inhomogeneous on the attractor. The variation of the largest Lyapunov exponents is seen to be relatively small, and the average over the variation shown in Fig. 5.33 for each field is $LE \approx 0.122/\text{h}$ or $LE \approx 1/(8.2 \text{ h})$. From that we estimate that predictions should be accurate, assuming we have accurately estimated the states at the beginning of predictions, up to the order of 10 or 20 h after the start of predictions.

The positive value for the largest Lyapunov exponent is the signal that the flow is chaotic (Abarbanel 1996; Kantz and Schreiber 2004).

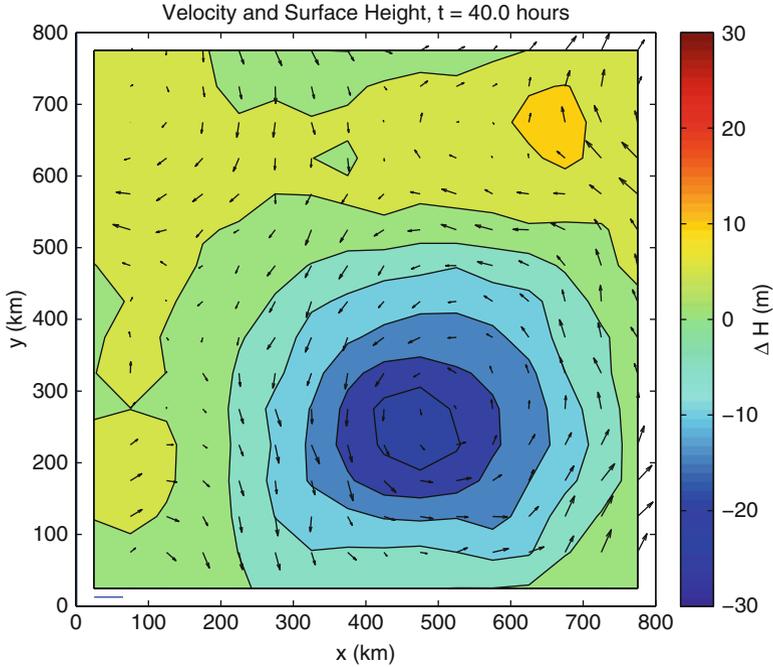


Fig. 5.32 Display of $h(i, j, t = 40 \text{ h})$ and $(u(i, j, 40 \text{ h}), v(i, j, 40 \text{ h}))$ from initial value integration of a 16×16 one-layer shallow water equation driven by surface Ekman pumping. The integration was carried forward $\approx 2,000$ simulated hours with a time step of 0.01 h . This is the height and horizontal velocity fields at the end of a 40 h observation period following that initial transient flow

5.5.4 Synchronization of the Data with the Model Output

Our model is now employed to generate the synchronization error for each of the fields $\phi_{i,j}(t) = \{u(i, j, t), v(i, j, t), h(i, j, t)\}$ by evaluating

$$\text{RMS synchronization error} = \frac{1}{T_{\max}} \sum_{t=0}^{T_{\max}} \left\{ \sum_{\text{observed } i,j} (\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))^2 \right\} \quad (5.50)$$

as a function of T_{\max} for various configurations of the observed set of grid locations (i, j) . As a notational remark, we use the integer labels (i, j) to indicate locations on the grid, while the independent variables (x, y) indicate continuous spatial field labels. We want to know how many and which measurements are required to synchronize the data and the model output. The number of data streams that must be observed we call L_c . Which L_c observations are required will be, as we will soon see, a matter of searching among the possibilities.

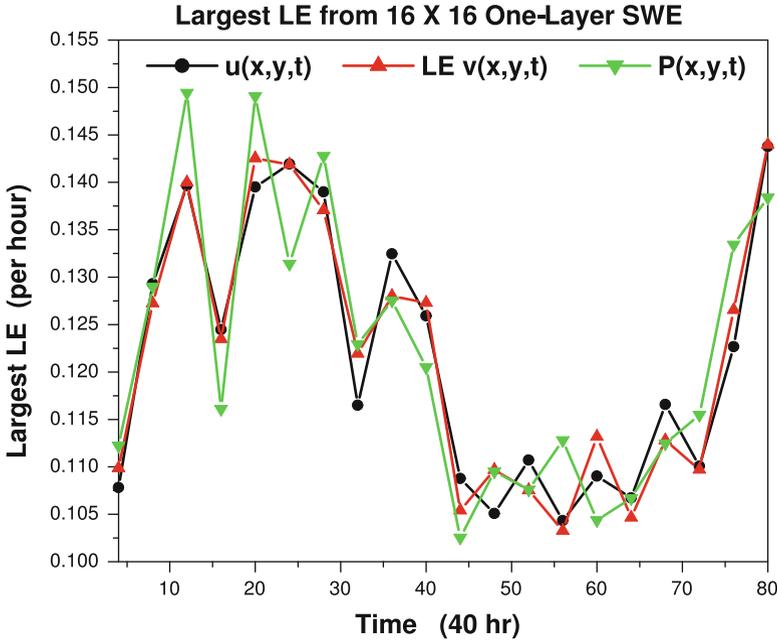


Fig. 5.33 The largest Lyapunov exponents for the 16×16 Ekman pumped one-layer shallow water equations. Two time series for the fields $\{h(\mathbf{r}, t), u(\mathbf{r}, t), v(\mathbf{r}, t)\}$ were evaluated with a small change in initial conditions for the two time series. The growth in the RMS errors for the difference in each field over 40 h segments of time were tracked and fit to $\exp[LE(t/1\text{ h})]$ in each time block. The exponent “LE” is the largest Lyapunov exponent for each field, and this varies slightly over different segments of the attractor as shown. The average for each field over the 80 blocks of length 40 h is 0.122/h or 1/(8.2 h)

To couple the information from the observations into the model, we add a term $k(\phi_{ij-\text{data}}(t) - \phi_{i,j-\text{model}}(t)); k \geq 0$ to the right-hand side of the differential equations for $\phi_{ij}(t)$ that correspond to an observed field. For example, if we observe $h_{1,2-\text{data}}(t)$ then in the differential equation for $h_{1,2-\text{model}}(t)$, we add the term $k(h_{1,2-\text{data}}(t) - h_{1,2-\text{model}}(t))$. In this calculation all observations are coupled into the model equations with the same value for the constant k .

With different initial conditions for $\phi_{i,j-\text{model}}(0)$ than were used in generating the data $\phi_{i,j-\text{data}}(t)$, we evaluate the RMS synchronization error for each field as a function of T_{max} for several values of k . In this we found that when k is larger than approximately 0.25 synchronization appears. This means that as T_{max} increases, for $k \geq \approx 0.25$ the RMS synchronization error goes to very small values. This can also be seen by choosing T_{max} large and fixed and then evaluating the sum for different values of k . Now using $k = 0.4$ for the first 40 h of integration, we can examine various selections of L_c and observation locations numbering L_c to see if synchronization occurs. For a second 40 h of integration, we reset k to 0. $k = 0$ is the autonomous chaotic flow.

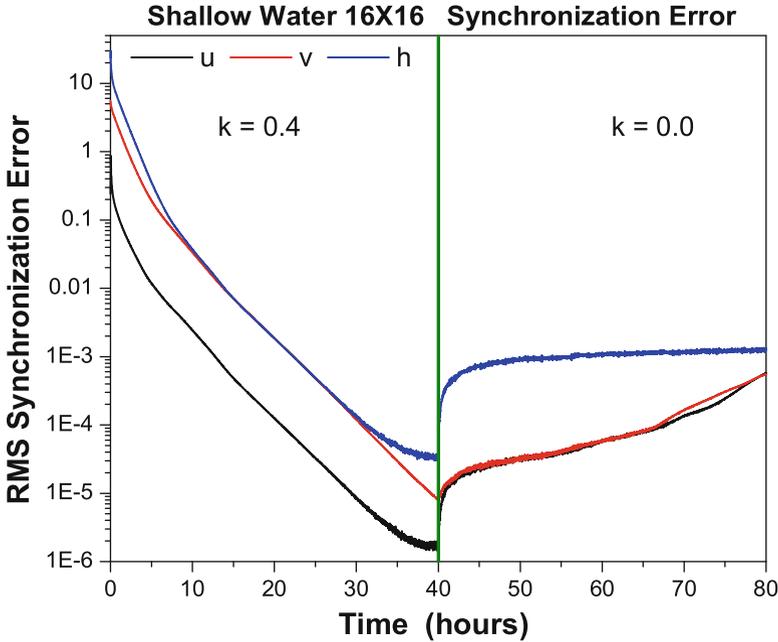


Fig. 5.34 The RMS synchronization error $\text{RMS synchronization error} = \frac{1}{T_{\max}} \sum_{t=0}^{T_{\max}} \sum_{\text{observed } i,j} (\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))^2$ between two chaotic trajectories of the 16×16 , Ekman pumped one-layer shallow water equations. In the first 40 h of calculation with time step 0.01 h, the data was coupled via $k(\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))$ into the differential equation for each field in the model receiving information from the data. The coupling was $k = 0.4$ in this segment. k was set to zero for the remaining 40 h of the calculation. We see that the RMS errors, Eq. (5.50), between the two time series for $t > 40$ h increase as expected for chaotic trajectories. In this calculation data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $u(i, j, t)$, and the 16 grid points of the Northernmost row of $v(i, j, t)$ were presented to the model equations. This shows synchronization between the two time series is achieved with this selection of 528 (69%) of the 768 independent fields of the total degrees of freedom

Since the data and the model output are chaotic, we expect the two trajectories in the second time segment of 40 h to move apart across the attractor as they arise from different initial conditions. The decorrelation between the trajectories associated with their separation on the shallow water attractor increases the synchronization error until it saturates near the size of the attractor.

In Fig. 5.34 we display the RMS synchronization error for each of the fields $\phi(i, j, t)$. In this calculation data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $u(i, j, t)$ and the 16 grid points of the Northernmost row of $v(i, j, t)$ were presented to the model equations. For the first 40 h of integration $k = 0.4$, and we couple the observed fields into the shallow water equations. We see that the synchronization error for each of the fields u, v , and h decreases to order 10^{-4} or smaller indicating the data and the model output are synchronized. After we turn off the coupling at $t = 40$ h, so $k = 0$, the model is evolving

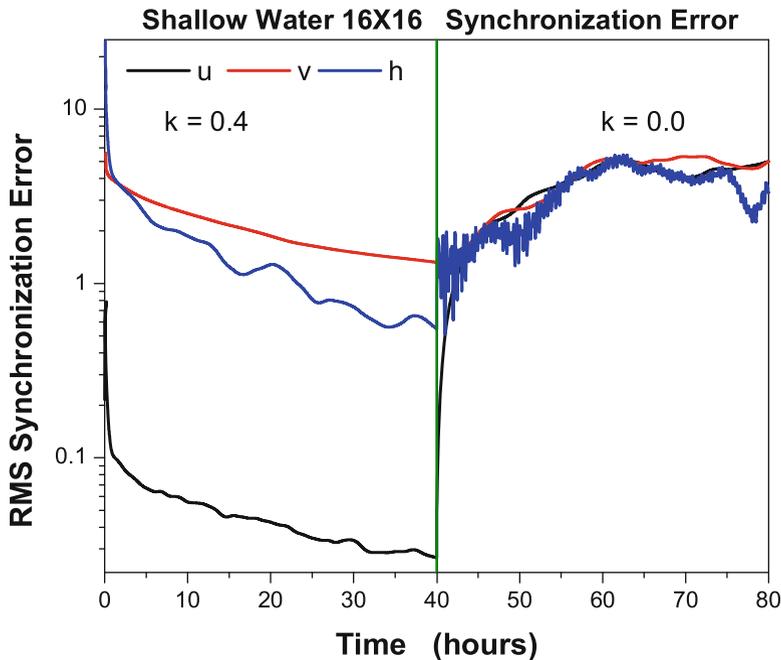


Fig. 5.35 The RMS synchronization error $\text{RMS synchronization error} = \frac{1}{T_{\max}} \sum_{t=0}^{T_{\max}} \sum_{\text{observed } i,j} (\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))^2$ between two chaotic trajectories of the 16×16 , Ekman pumped one-layer shallow water equations. In the first 40 h of calculation with time step 0.01 h, the data was coupled via $k(\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))$ into the differential equation for each field in the model receiving information from the data. The coupling was $k = 0.4$ in this segment. k was set to zero for the remaining 40 h of the calculation. We see that the RMS errors, Eq. (5.50), between the two time series now increase as expected for chaotic trajectories. In this calculation data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $u(i, j, t)$, and the 16 grid points of the westernmost column of $v(i, j, t)$ were presented to the model equations. This shows no synchronization between the two time series

without information from the data; the chaotic model output and the chaotic data become unsynchronized. This happens because at $t = 40$ h, the values of the model state variables, though possibly very close to the data values, have residual small errors when compared to the data, and this error grows as $\exp[\text{LE}(t/1 \text{ h})]$ because each is chaotic. This absence of synchronization is associated with positive CLEs on the synchronization manifold (Abarbanel 1996; Kantz and Schreiber 2004) where $\text{data}(t) = \text{model output}(t)$. The surfaces over which one must search for unobserved state variables or unknown parameters is replete with local minima when $k = 0$ and is smooth and easy to search for $k = 0.4$. In further examples below, we will see that it is not only the number of observed state variables that matters, but it also matters which ones are observed.

In Fig. 5.35 we perform the same calculation as before with data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $u(i, j, t)$, and the 16 grid points

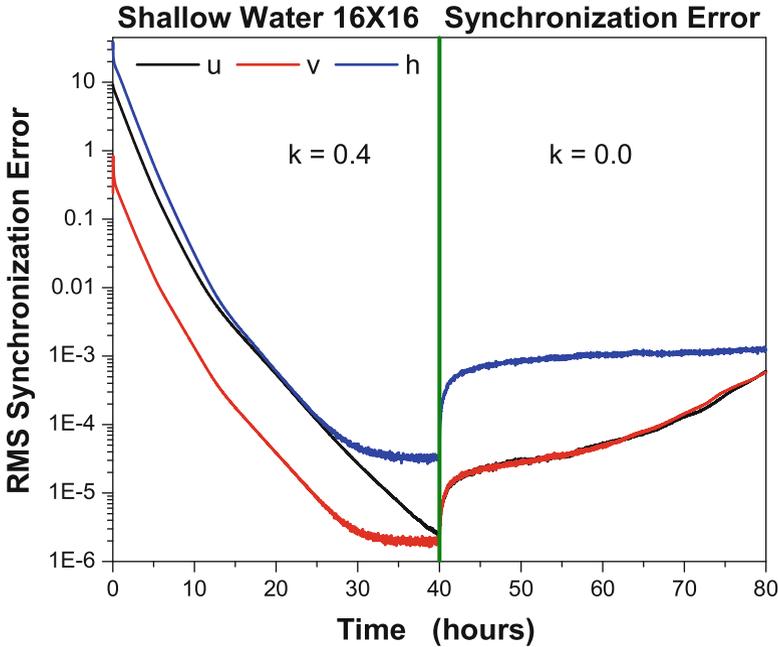


Fig. 5.36 The RMS synchronization error $\text{RMS Synchronization Error} = \frac{1}{T_{\max}} \sum_{t=0}^{T_{\max}} \sum_{\text{observed } i,j} (\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))^2$ between two chaotic trajectories of the 16×16 , Ekman pumped one layer shallow water equations. In the first 40 h of calculation with time step 0.01 h, the data was coupled via $k(\phi_{i,j-\text{data}}(t) - \phi_{i,j-\text{model}}(t))$ into the differential equation for each field in the model receiving information from the data. The coupling was $k = 0.4$ in this segment. k was set to 0 for the remaining 40 h of the calculation. We see that the RMS errors, Eq. (5.50), between the two time series now increase as expected for chaotic trajectories. In this calculation data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $v(i, j, t)$ and the 16 grid points of the Westernmost column of $u(i, j, t)$ were presented to the model equations. This shows synchronization between the two time series

of the Westernmost column of $v(i, j, t)$ presented to the model equations. In this instance, we see that synchronization did **not** transpire. Finally in Fig. 5.36 we once again perform the same calculation now using data from all 256 grid points for $h(i, j, t)$, all 256 grid points for $v(i, j, t)$, and the 16 grid points of the Westernmost column of $u(i, j, t)$ presented to the model equations. In this third setting we see that synchronization of the data and the model output again occurs. In each of the three examples, $L_c = 528$; however, we find that in one of the choices of L_c observations we do not achieve synchronization. We now turn to the ability to predict beyond a 40 h observation period for these selections.

5.5.5 Results for the Shallow Water Equations: Synchronization Implies Predictability

Using the synchronization error calculations just outlined, we identified selections of observations among the $\{u(i, j, t), v(i, j, t), h(i, j, t)\}$ where we see synchronization on the $\mathbf{y}(n) \approx \mathbf{x}(n)$ synchronization manifold as well as some selections of observations where synchronization does not occur. In each case we use Metropolis-Hastings Monte Carlo methods to evaluate the path integral in order to estimate the mean trajectory and the RMS variation about the mean from the beginning of an observation window for 40 h with a time step $\Delta t = 0.01$ h or 3.6 min. The action was approximated as coming from Gaussian noise in the measurements and a Gaussian approximation to the transition probability for the model dynamics.

In our calculations we used 200,000 updates of the path. The first 100,000 updates were excluded from the statistics on the expectation values. To broaden the basin in which the estimation procedure searched, we used the device of simulation annealing wherein the measurement error and model error are scaled by a common factor β_0 acting like an inverse temperature, then at selected times in the sequence of Monte Carlo iterations, β_0 is increased—that is, the effective temperature is decreased, until we arrive at $\beta_0 = 1$. Simulated annealing was done over the first 80,000 iterations, with various values for the rate at which a simulated annealing factor β_0 was introduced. The targeted acceptance rate was $f_{\text{acc}} = 0.4$, and the step size adjustment rate was $\alpha = 0.1$. The initial settings for Monte Carlo step sizes were $\Delta_i = 0.1$ for $U(i, j)$, $V(i, j)$ and $\Delta_i = 0.875$ for the $P(i, j)$ state variables. The scaling factor of 8.75 was used because the dynamical range of $P(i, j, t)$ was about 8.75 times larger than the dynamical range of $U(i, j, t)$ and $V(i, j, t)$.

We used $m = 4,000$ data points or 40 h of observations for the statistical data assimilation. The integral estimated using Monte Carlo methods has dimension about 2.8×10^6 . The matrix \mathbf{R}_f was taken to be diagonal with 12.8 on the diagonal for the $U(i, j)$, $V(i, j)$ entries and 0.17 on the diagonal for the $P(i, j)$ entries. The matrix \mathbf{R}_m was also set to be a diagonal matrix with 100.0 on the diagonal for the observed $U(i, j)$, $V(i, j)$ entries and 1.0 for the observed $P(i, j)$ entries. $R_m(i, j)$ is 0 for the times where no observations were made.

Each run of 200,000 iterations took 12.3 h on an NVIDIA GTX 580 which has 512 CUDA threads. The whole process was repeated five times, for a total of a million iterations, each time starting off with the final path from the previous run. The five runs differ by the initial cooling value β_0 . It was set to 0.0001, 0.1, 0.1, 0.1, 1.0 in sequence. The runs were repeated until the total action stopped decreasing significantly. Only the last 100,000 iterations of the final run were used in evaluating the statistics.

Ten different observation configurations were all run simultaneously on a CPU/GPU cluster running the Rocks Cluster Distribution software and the Oracle Grid Engine queuing system. The cluster is made up of three compute nodes, each with four NVIDIA GTX 580 GPUs installed, and a head node with two additional GTX 580s and an interface to the user and a high bandwidth Infineon

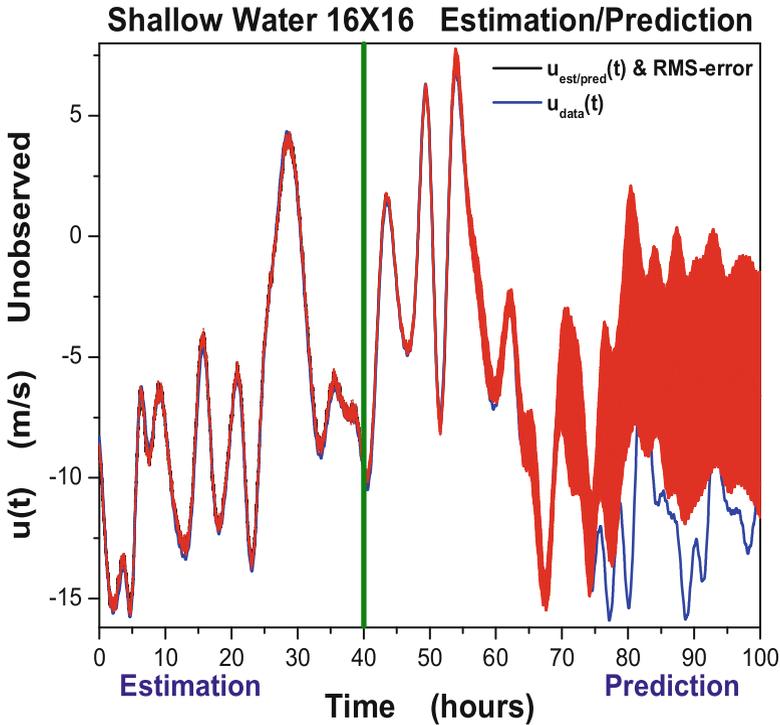


Fig. 5.37 A display of the comparison between the known value of an **unobserved** selected field $\phi_{i,j}(t)$ during the observation period of 40 h and during the prediction period of an additional 60 h. The expected value of an **unobserved** value of $u(i,j,t)$ along with the RMS error in the estimate for a case when all values of $v(i,j,t)$ and all values of $h(i,j,t)$ were observed and passed to the model of the Ekman pumped shallow water flow. The Westernmost column of values for $u(i,j,t)$ were also observed; the data and the model output **synchronize**. The estimates are accurate and the predictions are accurate for about 25–30 h after the end of the observation period. This is consistent with the rate of divergence of the chaotic trajectories according to our estimate of the largest Lyapunov exponents for this flow. There are no observations in the prediction window

communications node. Each of the configurations of observations was assigned to a single GPU and was run independently.

We selected different subsets of the 768 fields $\phi_{i,j}(t)$ to evaluate the ability of the path integral method to accurately estimate the unobserved states during and at the termination of the observation window at $t = T = 40$ h. Predictions were made by using all of the accepted paths during the observation window as initial conditions for the prediction from $t = T = 40$ h to $t = 100$ h. The mean prediction and the RMS variation about that mean are shown in each of the following cases. In the Left Panel for each selection of observed quantities, we show the expected values along with the RMS errors about the mean, and in the Right Panel we display the same quantities but remove the error bars for clarity of the display. The fixed parameters in the shallow water equations were not estimated in these calculations.

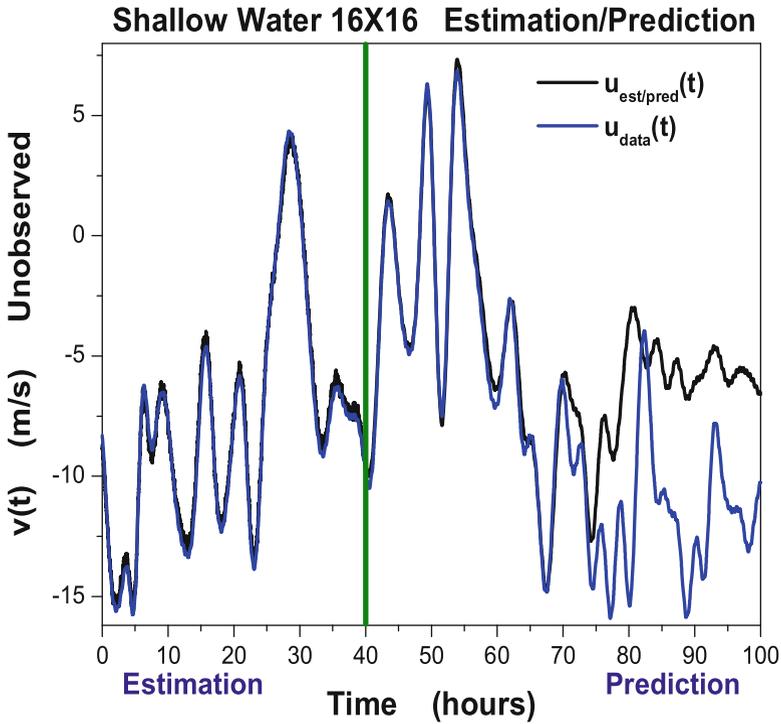


Fig. 5.38 The same calculation as in Fig. 5.37 without the error bars displayed. Removing these is for convenience in viewing the expected values. There are no observations in the prediction window

Our first example uses observations at each value of $v(i, j, t)$ and each value of $h(i, j, t)$ and the Westernmost column of $u(i, j, t)$ during the observation period. **For this set of observations the data and the model output synchronize.** Removing any of the observations from this set leads to a lack of synchronization. The results are displayed in Figs. 5.37 and 5.38. We show the time series for an observed value of $u(i, j, t)$ both in the period $0 \leq t \leq 40$ h and for the prediction window $40 \leq t \leq 100$ h. The estimates are quite accurate, and we conclude from the accuracy of the predictions that all unobserved state variables were accurately estimated using the path integral method. The predictions are accurate for about twice the inverse largest Lyapunov exponent which is consistent with the chaotic behavior of the dynamics and the estimation errors at the end of the observation window. An error at time T of $|\Delta\mathbf{x}(T)|$ grows to the size of the attractor, which we call one here, as $|\Delta\mathbf{x}(t + T)| \approx |\Delta\mathbf{x}(T)| \exp[\text{LE}(t/1 \text{ h})]$. We reported above that $\text{LE} \approx 1/8.2$ h, and our estimates at T are accurate to a few percent of the known value.

Our second example uses observations of all values of $u(i, j, t)$ and all values of $h(i, j, t)$ along with the Northernmost row of values of $v(i, j, t)$. The results are

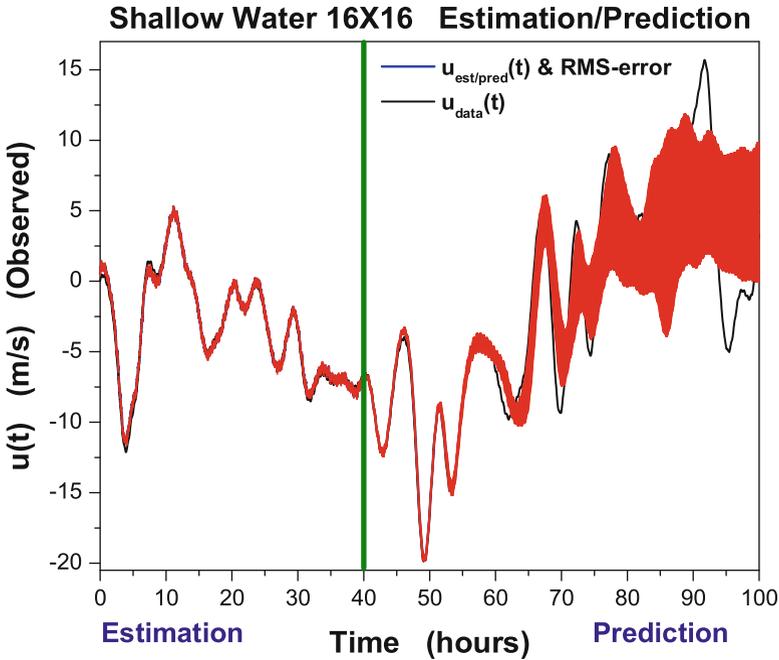


Fig. 5.39 A display of the expected value of an **observed** value of $u(i, j, t)$ along with the RMS error in the estimate for a case when all values of $u(i, j, t)$ and all values of $h(i, j, t)$ were observed and passed to the model of the Ekman pumped shallow water flow. The Northernmost row of values for $v(i, j, t)$ was also observed. Synchronization of the data and the model occurs. The estimates are accurate and the predictions are accurate for about 25–30 h after the end of the observation period. This is consistent with the rate of divergence of the chaotic trajectories according to our estimate of the largest Lyapunov exponents for this flow

shown in Figs. 5.39 and 5.40. In this example we observe all values of $u(i, j, t)$, all values of the height $h(i, j, t)$, and the Northernmost row of the y -velocities, $v(i, j, t)$. **This is a situation where we have synchronization between the data and the model output.** Indeed, as we expect by now, we see that the estimation in this case is quite good, both for the displayed $u(i, j, t)$ and, as evidenced by the good prediction, for all the unobserved values of $v(i, j, t)$.

The third example we present here is an unobserved value of $u(i, j, t)$ in the case where we have observed all values of $v(i, j, t)$ and all values of $h(i, j, t)$, but no values of $u(i, j, t)$. **This is a situation where we do not have synchronization of the data with the model output,** and we are able to see this both in the observation period because this is a twin experiment and in the prediction period where we can infer from the inadequate prediction that the unobserved states were not well estimated. In a realistic situation where we do not have knowledge of the “unobserved” state variables, that is, not a twin experiment, we could not know that the unobserved variable we selected was badly estimated in the prediction window $t \geq 40$ h. However, as we know there is no synchronization between the model and

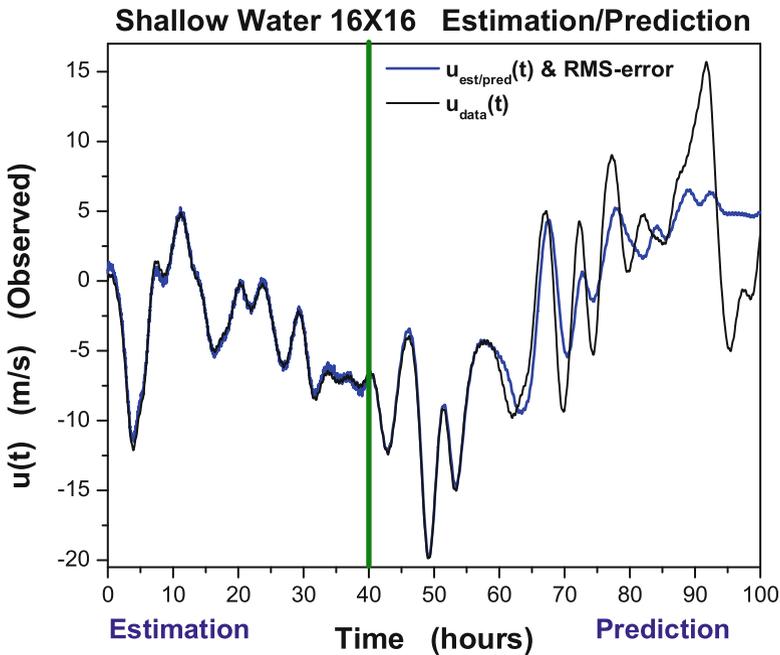


Fig. 5.40 The same quantities as in Fig. 5.39 without the error bars in order to allow a clearer view of the divergence of the trajectories in the prediction window associated with the chaotic behavior of the data and the model orbits. There are no observations in the prediction window

itself for this observation scenario, we would expect bad predictions as well. This result is shown in Figs. 5.41 and 5.42.

Our penultimate example is displayed in Figs. 5.43 and 5.44 is similar to our second example. Here we display an unobserved value of $v(i, j, t)$ when the observations, all $u(i, j, t)$, all $h(i, j, t)$, and the Northernmost row of $v(i, j, t)$, are sufficient to allow synchronization between the data and the model output. **As we have synchronization**, we expect accurate estimates of the unobserved states at the end of observations $T = 40$ h here, and from that knowledge of the full set of state variables at T , we see that good predictions ensue. The chaotic behavior of the orbits exhibits itself again by the divergence of the known (data) and the predicted values of this variable after about 25–30 h after the end of observations.

The final example we display has another message. Here we have observations of all values of $h(i, j, t)$ and various values of $u(i, j, t)$ and $v(i, j, t)$ but **not sufficient or appropriate observations to yield synchronization** between the data and the model output. However, as we see in Figs. 5.45 and 5.46 we have excellent estimates of the observed $u(i, j, t)$ variable in the observation window, nonetheless, we see quite inadequate predictions after the observations cease. This is the result of the inability of the data and the model output to synchronize and efficiently pass information from the data to the model. Estimations of the unobserved state variables apparently fail in this circumstance.

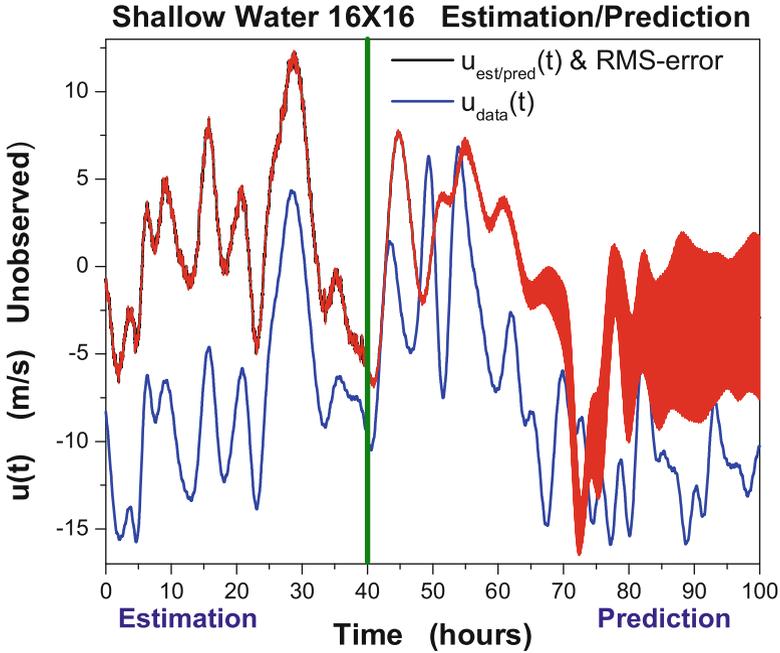


Fig. 5.41 A display of the expected value of an **unobserved** value of $u(i, j, t)$ along with the RMS error in the estimate for a case when all values of $v(i, j, t)$ and all values of $h(i, j, t)$ were observed and passed to the model of the Ekman pumped shallow water flow. No values of $u(i, j, t)$ were observed. There is no synchronization of the data and the model output. Neither the estimates nor the predictions are accurate with only these 512 fields observed. Display of the known values of the unobserved $u(i, j, t)$ along with the error bars in the estimation. The first 40 h is the observation window. The second 60 h is the prediction window. There are no observations in the prediction window

5.6 Synopsis and Perspectives: Twin Experiments

Twin experiments use no data but allow us to explore important issues about proposed physical relevant to extracting information about model states and parameters when data becomes available.

Addressing the number of measurements required to permit an accurate search for parameters and unobserved states is an important use of twin experiments. This is accomplished through examining synchronization of the model with data generated by the model itself when $1, 2, \dots, L$ measured quantities are available. Further, using twin experiments one can explore which measurements should be used.

When twin experiments reveal that the measurements required are not in the toolkit of the laboratory providing the data to inform and test one's model, conversations should ensue on how to address that deficiency. This is a move away

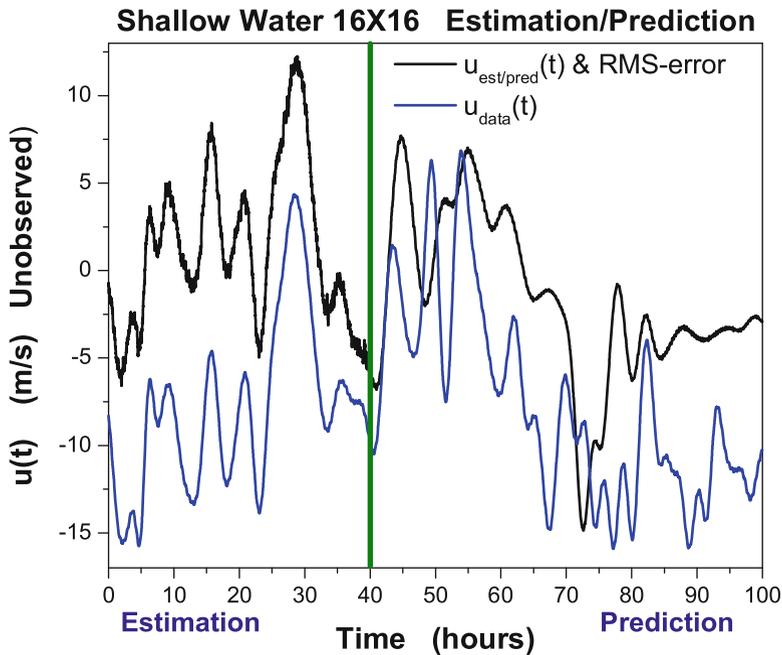


Fig. 5.42 The same quantities as in Fig. 5.41 without the error bars in order to allow a clearer view of the divergence of the trajectories in the prediction window associated with the chaotic behavior of the data and the model orbits. There are no observations in the prediction window

from a traditional view which provides data not particularly informed by the class of models one wishes to investigate and that can lead to dead ends in many cases.

A familiar use of twin experiments is to explore new methods for performing statistical data assimilation. This value remains.

If one can pass along advice to the reader, ours would be to use twin experiments for a thorough examination of the model or models one has in mind. Answering the kinds of question touched on in this chapter puts one in a good position to understand when your models, presented with laboratory or field data, are shown to be inconsistent with the data, where the inconsistency arises. In the more favorable circumstance when the model is consistent with the data and predicts model responses to additional forcing, one is in a good position to understand the dynamics of that success.

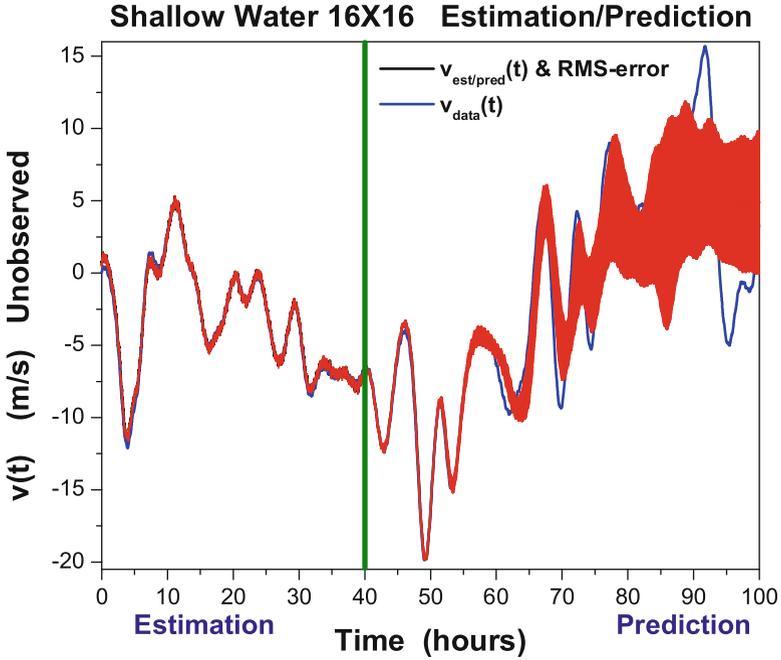


Fig. 5.43 A display of the expected value of an unobserved value of $v(i, j, t)$ along with its RMS error in the estimate for a case when all values of $u(i, j, t)$ and all values of $h(i, j, t)$ were observed and passed to the model of the Ekman pumped shallow water flow. The Northernmost row of values for $v(i, j, t)$ were observed, and synchronization of the data and the model does occur. The estimates are accurate and the predictions are accurate for about 25–30 h after the end of the observation period. This is consistent with the rate of divergence of the chaotic trajectories according to our estimate of the largest Lyapunov exponents for this flow. Display of the known values of the unobserved $v(i, j, t)$ along with the error bars in the estimation. The first 40 h is the observation window. The second 60 h is the prediction window. There are no observations in the prediction window

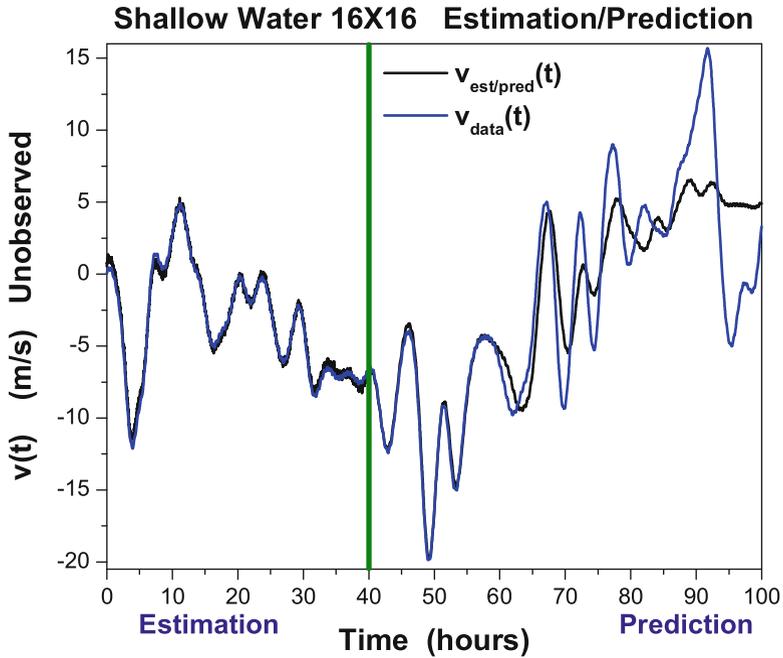


Fig. 5.44 The same quantities as in Fig. 5.43 but without the error bars in order to allow a clearer view of the divergence of the trajectories in the prediction window associated with the chaotic behavior of the data and the model orbits. There are no observations in the prediction window

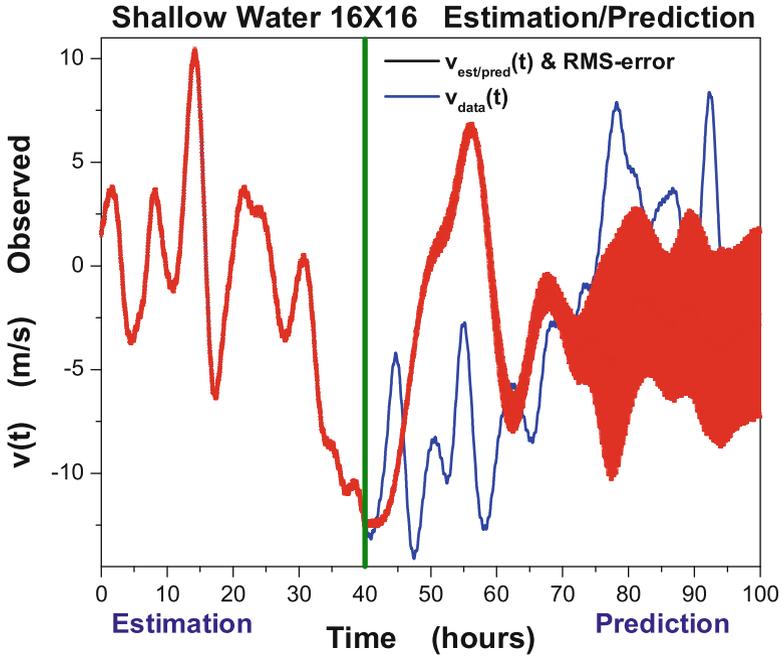


Fig. 5.45 A display of the expected value of an observed value of $v(i, j, t)$ along with the RMS error in the estimate for a case when all values of $h(i, j, t)$ and various values of $v(i, j, t)$ and $u(i, j, t)$ were observed and passed to the model of the Ekman pumped shallow water flow. Synchronization of the data and the model does not occur. The estimates are accurate but the predictions are not. We display of the known values of the unobserved $v(i, j, t)$ along with the error bars in the estimation. The first 40 h is the observation window. The second 60 h is the prediction window. There are no observations in the prediction window

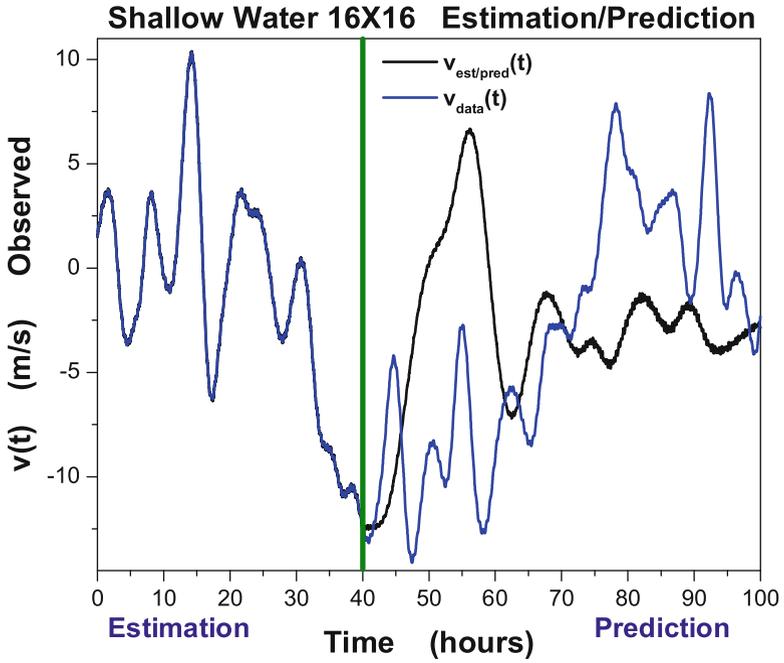


Fig. 5.46 The same quantities as in Fig. 5.45 but without the error bars in order to allow a clearer view of the divergence of the trajectories in the prediction window associated with the chaotic behavior of the data and the model orbits. There are no observations in the prediction window

Chapter 6

Analysis of Experimental Data

Prologue to the Chapter

All but two of the examples we have discussed in this book are twin experiments where laboratory or field data is not available. The example of the Colpitts circuit (Quinn et al. 2009) seen in Chap. 2 was a mixture of simulation and analysis of experimental data. Also in the example of the Malkus waterwheel (Illing et al. 2012a), experimental data was available, but we did not use it.

The significant challenges of working with laboratory data include (1) noise in the measurements and (2) uncertainty about the ingredients of quantitative models of the physics underlying the observations. Indeed, one important aspect of the statistical data assimilation procedures we have described can be viewed as tools for determining the consistency of proposed models of the observations.

The remainder of this chapter is devoted to a somewhat detailed discussion of the acquisition and analysis of experimental data sets from a neurobiology laboratory. This is a serendipitous possibility because of the author's connection with the laboratory of Daniel Margoliash at the University of Chicago and the laboratory of Timothy Gentner at UC, San Diego. Three important advantages of working with these laboratories are that precise, insightful experiments can be performed in them, the leaders of the laboratories are keenly interested in the outcomes of data assimilation analyses for the development of their own fields, and finally a set of basic neurobiological equations, the Hodgkin–Huxley equations (Johnston and Wu 1995; Koch 1999), are available as a quantitative starting point for model making.

These advantages may not yet be always fully available in other areas of scientific inquiry, so we take advantage of this benefit now, expecting the reach of the analysis tools and methods will be longer in the future.

The measurements on individual neurons presented here come from the neural substrate of the avian song system. After giving a rapid introduction to the issues within that biological system, we turn to stimulus/response experiments performed on neurons extracted from a critical nucleus in the birdsong network. It is important to dwell on some of the details of the measurements and the experimental

preparation, and this is done. The full interpretation for biological understanding of the birdsong system is left to (Meliza et al. 2013), but the reader should recognize many of the important issues in working with noisy data and uncertain models in a realistic and scientifically important setting.

6.1 The Avian Song System: Individual Neurons

Nervous systems are comprised of a few to very many individual nonlinear oscillators, the neurons, connected by a variety of chemical and electrical links. They differ from the improvised networks often studied as complex systems as they have evolved to perform functions as have other networks in living systems. Some biological networks are simple to describe, their function is simple to state, and their description in biophysical terms is involved. One of these, the pyloric central pattern generator (Stein et al. 1997) found in many crustaceans, has only fourteen neurons and its function is to guide a muscle chamber through a three-phase motion used to move shredded food to the animal's digestive system. It is certainly a complex system, hardly in league with the mammalian brain, yet as one of the very best characterized networks of neurons, it has been both analyzed in depth and used as a model for testing ideas about neural networks (Elson et al. 1998).

Of intermediate size is the avian song system which may be one of the smallest networks which learn a function, vocalization in this case, from a previous generation and then use it for functional needs. Here the song is utilized for attracting mates and defending territory. The song is then passed along to another generation (Zeigler and Marler 2004). The entire song system comprises perhaps a collection of a million or so neurons organized in nuclei each of order 10^4 or 10^5 neurons forming pathways dedicated to producing song, learning the song from an older male, maintaining the song, memorizing the song, and regulating the song.

Over fifty years much has been learned about the construction and function of this network, and our interest here is the analysis of the detailed biophysics of the circuits involved starting with the characterization of one of the key nuclei, HVC. This we do with experimental data on individual neurons from the three general classes of neurons in HVC. If this is successful, using the neuron models established by this analysis, one would seek to construct a model of the network links that enable the classes of neurons to exercise the functions required of HVC. Much is known about those functions both in a qualitative and in a quantitative manner, and the results of these excellent experiments constrain the quantitative, predictive models one seeks to develop.

We turn now to the analysis of some laboratory neurobiology experiments on the birdsong system performed by C. Daniel Meliza in the laboratory of Daniel Margoliash at the University of Chicago. We will address only a very small subset of the data collected in these experiments as our purpose here is primarily illustrative rather than an exhaustive exposition of the biological issues illuminated by the data collection protocols and the analysis of the biophysics of the neurons involved (Meliza et al. 2013).

The experiments were performed on individual neurons from an important segment of the avian songbird brain. The song system nervous network is found in brain areas “above” the brain stem. The brain stem neural activity directly controls the musculature of the songbox, also known as the syrinx (Laje and Mindlin 2005). It is a more-or-less cylindrical column branched to the two lungs. Voltage stimulation changes the tension of the muscles and the resulting frequency or pitch changes. Expiration and inspiration involving the lungs create the vocalizations we call song.

The neural apparatus shapes the signals to the syrinx and the lungs. It is responsible for the instruction sets to create the vocalizations, for the memory repository to remember the song from males of an earlier generation, for filtering the self-produced song for its fidelity with remembered song, for the maintenance and control of the song quality by internal neural feedback loops, and for the external/internal auditory feedback from the bird’s beak to a central nucleus in the brain named HVC. Juvenile birds have a genetically controlled vocal apparatus which, species by species, appears to be generally tuned to the song of its species. In its early sensory period the juvenile hears song from males in its vicinity and remembers what is called a “template” of that song. At about 30 days of age the juvenile begins singing to itself both audibly as produced vocalizations and via neural activity in sleep periods. During this sensory-motor period, projections from the HVC are “wired up” via plasticity of synaptic connections.

The experiments in the Margoliash laboratory were performed on neurons from the HVC in the species zebra finch (*Taeniopygia guttata*) which appears to learn only one sequence of syllables, each about 100–150ms in length, to form a vocalization motif about one second in length. This song is maintained and used throughout the adult life of the finch.

We are very far from being able to represent in a quantitative model the full range of activity of the avian song system, though that is an ambitious goal. The experiments we discuss are performed on single, isolated neurons from the zebra finch HVC. The framework for studying individual neurons is that the nervous systems whose behavior we wish to understand are composed of the individual neurons at the nodes of a network. In order to build a model of these networks with a high level of biophysical realism, we start with the dynamics of the nodes.

However important the biophysics of individual neurons may eventually be in the discussion of the song system network, the connectivity among the nodes will play a key role in the functional activity of the song network. The analysis of the network connections would follow what we will address about individual neuron cells.

To use the methods developed in this book, we need a biophysical model and we need biophysical experimental data to complete that model by estimating the many parameters in the model. The biophysical modeling framework we adopt is that of the ion-current-based Hodgkin–Huxley (HH) models (Johnston and Wu 1995; Koch 1999) we have explored in many instances.

In designing the experiments and acquiring the data, we take many lessons from the twin experiments explored throughout this book. For example, we select our model to have a large number of intrinsic neuron currents driving the voltage

response of the neuron. In twin experiments we saw that if a current is absent in the data, it will be pruned from the model by our procedures. Of course, we may still miss currents not included in the large model; however, the selection made appears to work rather well.

The data we acquired in our experiments are the voltage $V(t)$ responses of individual neuron cells across the cell membrane. These are associated with the injection of an observer-selected time-dependent current $I_{\text{applied}}(t)$ into the cell. The HH model we select is certain to be “wrong” both in missing some aspect of the biophysical processes determining $V(t)$ in response to $I_{\text{applied}}(t)$ and in having to select a functional form for the HH model that is phenomenological and not derived from any first principles of biological physics. The data also has noise, and as only the state variable $V(t)$ is observable, we have little clue as to the full state of the model, never mind the actual neuron, when measurements commence.

6.2 Experimental Procedures for HVC Neurons

We work with slices of the avian brain taken from the nucleus HVC. In vivo experiments have the problem that the neurons whose properties we wish to estimate and test are linked to the other neurons in the intact functional network. To avoid having to represent those currents external to the neuron within our model, we worked with a slice preparation where individual identified neurons could be analyzed.

6.2.1 HVC Slice Preparation: Experimental Procedure

Slices were prepared from adult male zebra finches. Birds were deeply anesthetized with isoflurane, decapitated, and the brain placed in ice-cold, oxygenated dissection buffer. Parasagittal slices were cut ($400\ \mu$ thick, vibratome) from both hemispheres and placed in 37°C oxygenated artificial cerebrospinal fluid (ACSF). Slices were allowed to recover for at least 1 h, during which time the ACSF was allowed returned to room temperature. For recording, slices were superfused with oxygenated ACSF ($23\text{--}26^\circ$).

Current-clamp somatic whole-cell recordings were made from neurons in HVC with an NPI SEC-05L amplifier (NPI Electronic, Tamm, Germany). Cells were selected visually with differential IR optics for health but not for somatic size or shape. Patch pipettes ($3\text{--}5\ \text{M}\Omega$) were pulled from standard-walled borosilicate glass (Model G150F-4, Warner Instruments, Hamden CT) with a Sutter P-97 (Sutter Instruments, Novato, CA) and filled with an internal solution. Pipette capacitance and series resistance were compensated on the amplifier. Voltage and injected current (as reported by the amplifier) were high pass filtered at 20 kHz and digitized at 50 kHz with a PCI-6052E (National Instruments, Austin, TX). Data collection was controlled by custom MATLAB (The MathWorks, Natick, MA) software.

6.3 Experimental Results and Analysis

Data on the voltage response of individual neurons presented with specific injected currents were recorded and stored for analysis. Each neuron received a patch clamp electrode (Hamill et al. 1981) through which current was injected and at which voltage was measured every 0.02 ms. Neurons were individually held for some hours during which a selection of injected current protocols were presented for temporal epochs ranging from about two to about six seconds with rest periods between epochs. The estimation window consisted of 1,500 ms of current injection and observed voltage data. We recorded from some dozens of neurons with each having a few tens of measured epochs. Sampling at 50 kHz is five to ten times more rapid than is common in neuroscience; however, when performing twin experiments (Abarbanel et al. 2011) we found that the details of the voltage response, especially of subthreshold voltage variations, were missed at lower sampling rates. Estimates of parameters and states were also inaccurate at sampling rates below 20 kHz.

We used the path integral formulation whose solution was approximated by a saddle point (Laplace 1774) method with a HH model comprising many ion channels assumed to have no model errors. Only the stationary path was produced, though as usual all parameters in the model and all unobserved state variables were estimated. We chose a model with two Na channels, three K channels, two Ca channels, a “leak” (passive) channel, and an I_h channel. We anticipated that some of these channels might be absent in some of the neuron recordings; however, our strategy was formulated on the basis of the numerical experiments which showed that our procedures prune out ion channels in a model which are absent in the data presented to it. Of course, we do not know what channels are present in the data; that is one of the goals of the experiments. However, we found systematically that two of our selected currents were absent and others, for some neurons, were small indeed.

6.4 Model Details

Our biophysical model for the neurons is a single-compartment isopotential model with the membrane voltage V_m given by the current conservation equation

$$\frac{dV(t)}{dt} = \frac{1}{C_m} (I_{\text{NaT}}(t) + I_{\text{NaP}}(t) + I_{\text{K1}}(t) + I_{\text{K2}}(t) + I_{\text{K3}}(t) + I_h(t) + I_{\text{CaL}}(t) + I_{\text{CaT}}(t) + I_{\text{Leak}(t)} + I_{\text{inj}}(t)/I_{\text{SA}}), \quad (6.1)$$

where C_m is the specific membrane capacitance. I_{SA} is a parameter relating to the surface area of the membrane; it sets the scale of the injected current actually seen by the neuron. The I_X 's are channel currents. Each of the voltage-gated currents

Table 6.1 Currents used in the Hodgkin–Huxley model of HVC neurons

Ion current name	Ion current formulation
Transient sodium	$I_{\text{NaT}}(t) = g_{\text{NaT}}m(t)^3h(t)(E_{\text{Na}} - V(t))$
Persistent sodium	$I_{\text{NaP}}(t) = g_{\text{NaP}}m(t)(E_{\text{Na}} - V(t))$
Non-inactivating fast potassium	$I_{\text{K1}}(t) = g_{\text{K1}}m(t)^4(E_{\text{K}} - V(t))$
Inactivating potassium	$I_{\text{K2}}(t) = g_{\text{K2}}m(t)^4h(t)(E_{\text{K}} - V(t))$
Slow potassium	$I_{\text{K3}}(t) = g_{\text{K3}}m(t)(E_{\text{K}} - V(t))$
Hyperpolarization-activated cation	$I_{\text{h}}(t) = g_{\text{h}}h(t)(E_{\text{h}} - V(t))$
High-threshold L-type calcium	$I_{\text{CaL}}(t) = m(t)^2h(t)\text{GHK}([\text{Ca}]_{\text{intra}}, V(t))$
Low-threshold T-type calcium	$I_{\text{CaT}}(t) = m(t)^2\text{GHK}([\text{Ca}]_{\text{intra}}, V(t))$
Leak	$I_{\text{L}}(t) = g_{\text{L}}(E_{\text{L}} - V(t))$

depends on ion flow through channels whose permeability is controlled by activation $m(t)$ and inactivation $h(t)$ gating variables. For a given ion channel, except Ca^{2+} , the current is represented as $I_{\text{ion}}(t) = gm(t)^{n1}h(t)^{n2}(E_{\text{reversal}} - V_m(t))$, where g is a maximal conductance and $n1, n2$ are integers (Table 6.1).

For Ca^{2+} channels we replace the ohmic voltage dependence $E_{\text{reversal}} - V(t)$ by the GHK representation

$$\text{GHK}([\text{Ca}]_{\text{intra}}, V) = -V \frac{([\text{Ca}]_{\text{intra}} - [\text{Ca}]_{\text{extra}}e^{-2FV/RT})}{1 - e^{-2FV/RT}}, \quad (6.2)$$

where the intracellular and extracellular Ca^{2+} concentrations appear. The Faraday constant F and gas constant R are such that at body temperature, $\frac{RT}{F} = 26$ mV. Voltage-gated variables multiplying this GHK voltage dependence are for high-threshold L-type calcium currents (I_{CaL} ; $m(t)^2h(t)$) and for low-threshold T-type calcium currents (I_{CaT} ; $m(t)^2$).

The dynamics of the ion channel gating particles are given by voltage-dependent opening and closing rates. To ensure numerical stability, we use a hyperbolic tangent approximation to the Boltzmann barrier-hopping rate (Johnston and Wu 1995)

$$\begin{aligned} \frac{dm(t)}{dt} &= \frac{m_0(V(t)) - m(t)}{\tau_m(V(t))}, \quad m_0(V) = 0.5(1 + \tanh((V - V_{1/2})/\kappa)), \\ \tau_m(V) &= \tau_0 + \tau_{\text{max}}(1 - \tanh^2((V(t) - V_{1/2})/\sigma)), \end{aligned} \quad (6.3)$$

where $V_{1/2}$ is the half-activation voltage, κ is the slope of the activation function, τ_0 is the minimum relaxation time, $\tau_0 + \tau_{\text{max}}$ is the peak relaxation time, and σ is the width of the relaxation time function. Equations for the inactivation variables $h(t)$ have a similar form. The model has 12 time-varying state variables including $V(t)$ and 72 parameters, including the kinetic variables, maximal conductances, and reversal potentials for each of the channels, all of which were estimated during data assimilation.

The permeability of each voltage-gated channel species depends on one or more gating variables whose rate of opening and closing is modeled by its differential equation. For the K2 and CaT inactivation particles we used a more complex form that allowed relaxation of voltage dependence to be asymmetric. The membrane voltage along with the values for each of the gating variables forms a 12-dimensional vector $\{x_1(t), x_2(t), \dots, x_{12}(t)\}$ that describes the state of the neuron. The dynamics of the neuron are defined by a set of differential equations that depend on a number of unknown parameters. The complete set of model equations that are used for the optimization procedure, including the regularization term, are given here:

$$\begin{aligned} \frac{dx_1(t)}{dt} = \frac{1}{p_1} & \left\{ (p_2 x_2(t))^3 x_3(t) + p_3 x_4(t) (p_4 - x_1) + (p_5 x_5(t)^4 + p_6 x_6(t)^4) x_7(t) \right. \\ & + p_7 x_8(t) (p_8 - x_1(t)) + (p_{71} x_9(t)^2 + p_{72} x_{10}(t)^2 x_{11}(t)) \text{GHK}(p_{11}, x_1(t)) \\ & \left. + p_9 (p_{10} - x_1(t)) + p_{12} x_{12}(t) (E_h - x_1(t)) + I_{\text{inj}}(t) / p_{13} \right\} \\ & + u(t) (V_{\text{data}}(t) - x_1(t)) \end{aligned} \quad (6.4)$$

$$\begin{aligned} \frac{dx_2(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{14})/p_{15}) - 2x_2(t))/(p_{17} \\ & \quad + p_{18}(1 - \tanh^2((x_1(t) - p_{14})/p_{16}))) \\ \frac{dx_3(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{19})/p_{20}) - 2x_3(t))/(p_{22} \\ & \quad + p_{23}(1 - \tanh^2((x_1(t) - p_{19})/p_{21}))) \end{aligned} \quad (6.5)$$

$$\begin{aligned} \frac{dx_4(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{24})/p_{25}) - 2x_4(t))/(p_{27} \\ & \quad + p_{28}(1 - \tanh^2((x_1(t) - p_{24})/p_{26}))) \end{aligned}$$

$$\begin{aligned} \frac{dx_5(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{29})/p_{30}) - 2x_5(t))/(p_{32} \\ & \quad + p_{33}(1 - \tanh^2((x_1(t) - p_{29})/p_{31}))) \end{aligned}$$

$$\begin{aligned} \frac{dx_6(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{34})/p_{35}) - 2x_6(t))/(p_{37} \\ & \quad + p_{38}(1 - \tanh^2((x_1(t) - p_{34})/p_{36}))) \end{aligned} \quad (6.6)$$

$$\begin{aligned}
\frac{dx_7(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{39})/p_{40}) - 2x_7(t))/(p_{42} \\
&\quad + p_{44} + 0.5(1 - \tanh(x_1(t) - p_{39})) \\
&\quad \cdot (p_{43}(1 - \tanh^2((x_1(t) - p_{39})/p_{41})) - p_{44})) \\
\frac{dx_8(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{45})/p_{46}) - 2x_8(t))/(p_{48} \\
&\quad + p_{49}(1 - \tanh^2((x_1(t) - p_{45})/p_{47}))) \\
\frac{dx_9(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{50})/p_{51}) - 2x_9(t))/(p_{53} \\
&\quad + p_{54}(1 - \tanh^2((x_1(t) - p_{50})/p_{52}))) \\
\frac{dx_{10}(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{55})/p_{56}) - 2x_{10}(t))/(p_{58} \\
&\quad + p_{59}(1 - \tanh^2((x_1(t) - p_{55})/p_{57}))) \\
\frac{dx_{11}(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{60})/p_{61}) - 2x_{11}(t))/(p_{64} \\
&\quad + p_{65}(1 + \tanh((x_1(t) - p_{60})/p_{62})) \\
&\quad \cdot (1 - \tanh((x_1(t) - p_{60})/p_{63}))(1 - \tanh(x_1(t) - p_{60})) \\
&\quad \cdot \tanh((1/p_{62} + 1/p_{63})(x_1(t) - p_{60}))) \\
&\quad / (1 + \tanh((x_1(t) - p_{60})/p_{62}) \tanh((x_1(t) - p_{60})/p_{63}))) \\
\frac{dx_{12}(t)}{dt} &= 0.5(1 + \tanh((x_1(t) - p_{66})/p_{67}) - 2x_{12}(t))/(p_{69} \\
&\quad + p_{70}(1 - \tanh^2((x_1 - p_{66})/p_{68}))). \tag{6.7}
\end{aligned}$$

We do not search on the reversal potential for the I_h current, setting $E_h = -43$ mV, and we constrain E_{Na} , E_K , and E_L near $\{E_{Na} \approx 50$ mV, $E_K \approx -77$ mV, and $E_L \approx -65$ mV $\}$, respectively.

6.5 Details of the Numerical Evaluation

As we have suggested in several places in this book, one should check whether this model synchronizes with itself when only the voltage $V(t) = x_1(t)$ is passed from a data set created by the model to the model itself. This was checked, and as the structure of the Hodgkin–Huxley models is such that voltage drives all of the other state variables, synchronization occurs when we use parameters in a reasonable biophysical range.

We use the saddle-point approximation on the path integral, and we assume no model errors in this set of calculations; therefore, we have a constrained optimization problem as has been discussed extensively before. The optimization was accomplished using the open-source software IPOPT (Wächter and Biegler 2006) on standard desktop hardware. The data assimilation window over which the model properties are estimated was 1,500 milliseconds long. The data were sampled at 50 kHz, resulting in 75,000 time points of voltage data used in the observation window for each calculation.

Common to “direct method” variational approaches, each component of $\{x_1(t), x_2(t), \dots, x_{12}(t)\}$ at each observation time t_n was treated as an independent variable in the optimization procedure with the model dynamical equations imposed as equality constraints between neighboring time points. Gating particle variables were constrained between 0 and 1, and each of the parameter was constrained between biologically realistic bounds. The variational problem was 975,072 dimensional including the 75,000 values of the regularization variable $u(t)$ at the temporal co-location points.

The full or completed model, with estimated parameters and a full set of state variables at the end of the assimilation window $T = 1,500$ ms, was then integrated forward, with $u(t) = 0$, for the remainder of the data epoch with the same injected current that was presented to the real neuron. To generate predictions on other data epochs for the same neuron, a 100 ms long section of data was used to find the initial conditions for the forward integration of the state variables. The parameters were fixed at their previously established values in the original epoch. With parameters fixed in the model and initial conditions estimated in the new epoch, the model was integrated forward with the corresponding injected current and $u(t) = 0$.

We cannot overemphasize the importance of fully exercising the dynamical range of the neuron’s dynamics during the data assimilation window. This is achieved by subjecting the neuron to a current with a complicated waveform, exhibiting a broad power spectrum, thus having many characteristic timescales, as well as regions of constant (positive, negative, and zero) current to ascertain the neurons passive response properties. It is also quite important that the frequency content of the current be low enough as the neuron acts as a passive RC circuit low-pass filtering signals. If the frequency content of the applied current is too high, information in the current will be filtered out by the neuron itself.

6.5.1 *Estimation and Prediction of HVC Neuron Responses to Injected Current*

Whole-cell recordings were made from many HVC neurons stimulated by injection of various complex current waveforms designed to drive the cell through its range of biologically relevant states. The data assimilation procedure was *successful* for about two-dozen neurons. The metric for *success* is the ability of the completed model to predict the response of the neuron to current injections different from those used in the observation window [0 ms, 1,500 ms].

6.6 Results from Data Acquisition and Analysis

From a large database of stimulating currents and voltage responses from individual neurons in HVC, we have selected three to discuss. The neurons are known by the year and date and the temporal epoch as well as the order in which a particular neuron was studied during the given day. The names of the selected neurons—20110413.4.1, Epoch 22; 20120517.1.1, Epoch 11 and Epoch 12; and 20120406.1.3, Epoch 19 and Epoch 15—are not informative about the type of neuron within HVC: projection or interneuron.

Step currents are in common use among electrophysiologists and are interspersed within the complex waveforms used in our stimulation protocols. We checked that the models completed by the estimation of parameters in our HH models did respond to step currents as well as the actual neurons. The response to step currents is not as reliable as the neuron response to complex currents.

6.6.1 *Neuron 20110413.4.1 Epoch 22*

We show first the stimulating current and resulting voltage response for Epoch 22 of neuron 20110413.4.1; this is in Fig. 6.1. The injected current is a combination of a chaotic waveform, taken from the Malkus waterwheel equations (Malkus 1972), and a series of step currents with chaotic “decoration” on the constant current segments. This was selected on the criteria we have mentioned: all degrees of freedom of the model neuron are stimulated by this current, the frequencies in its Fourier transform are low enough not to be filtered out by the RC time constant of the neuron, and the stimulus is applied long enough to excite all currents contained in the dynamical model. We see that this neuron does not produce action potentials, spikes, very frequently. We do not know what class of neuron within HVC we have selected by our procedure of patch clamping neurons with no biophysical selection criterion, but indications come from further biophysical analysis (Meliza et al. 2013).

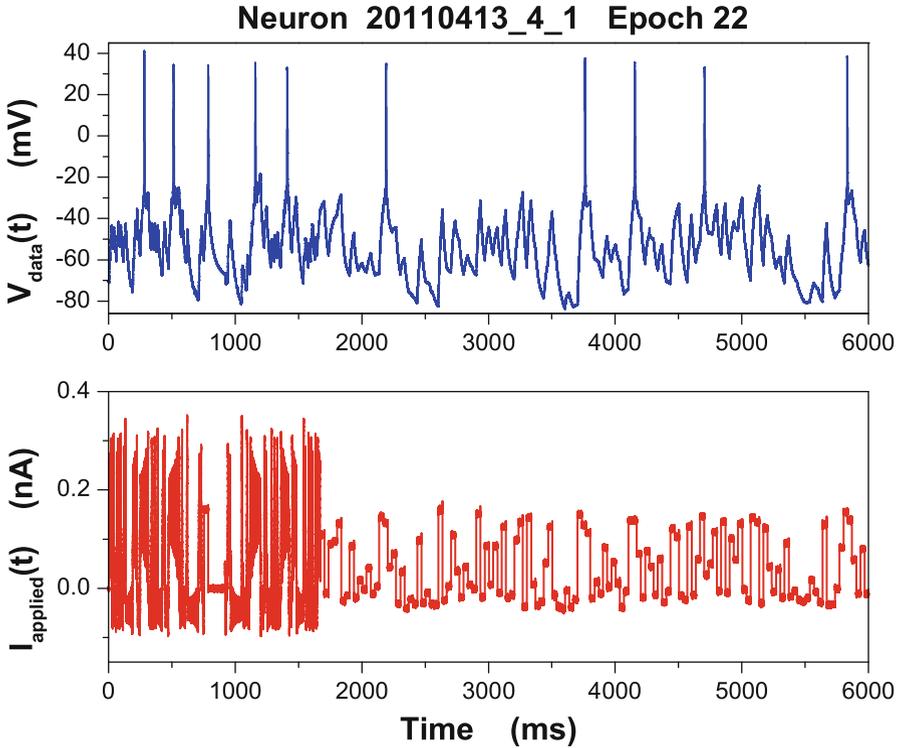


Fig. 6.1 Experimental data for neuron 20110413_4_1 during Epoch 22. Data was collected for six seconds in this epoch. *Top Panel* Membrane voltage response to the injected current shown in the *Bottom Panel*. The injected current is comprised of a complex waveform borrowed from a simple chaotic oscillator interspersed with a section of nearly zero current, then concluded with a sequence of nearly constant step currents with a small amplitude version of the chaotic waveform imposed. These two panels are characteristic of the data collected from the many neurons in the experimental efforts. The data for each epoch for each neuron comprises the forces imposed on the neuron by the injected current and the cross membrane voltage observed during an epoch. Estimated errors in the voltage measurements are ± 1 mV or less

We use $T = 1,500$ ms of the combined $V(t)$ and $I_{\text{applied}}(t)$ data to estimate the 72 parameters and all the gating variables over the time segment $[0 \text{ ms}, 1,500 \text{ ms}]$. The estimated time series for $V(t)$, the only observed state variable, is shown in red in Fig. 6.2. It is followed in the same figure with a prediction for $t > T$ using the estimated parameters in the differential equations along with the estimated values of $V(T)$ and all of the $m(T), h(T)$ for all the currents. This result is shown in blue. It clearly tracks the observed voltage data while producing one additional action potential near $t \approx 2,600$ ms. Prediction errors leading to additional or missing spikes are not uncommon in this procedure. The threshold for spike production, dynamically for moving the state of the neuron to the vicinity of the instability in state space leading to spiking, is quite sensitive to the particular values of parameters.

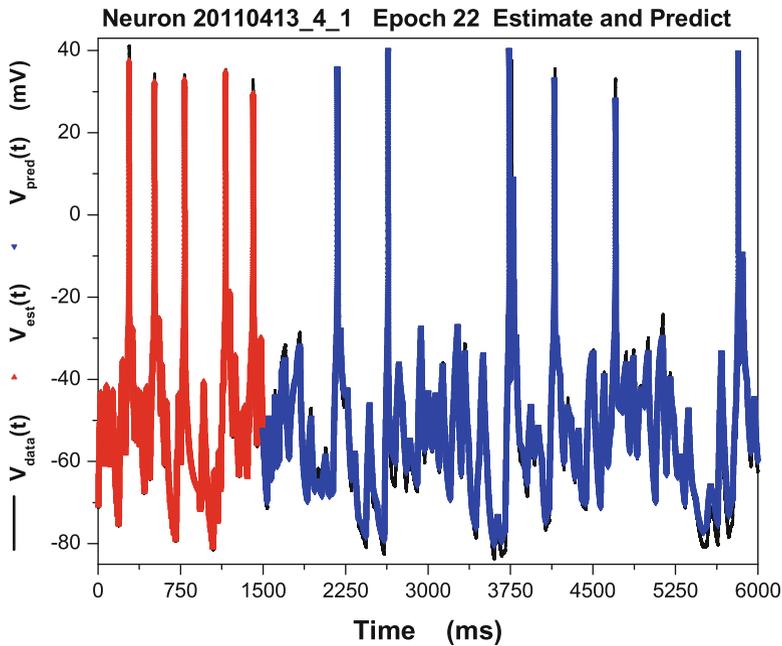


Fig. 6.2 Experimental data, estimated membrane voltage and predicted membrane voltage for neuron 20110413_4_1 during Epoch 22. The observed data is shown as a *heavy black line*. The estimated voltage during the observation window $[0, 1,500]$ ms is shown in *red*. It is during this observation window that the fixed parameters in the model are estimated and the eleven unobserved state variables, here the gating variables $m(t)$ and $h(t)$ for each ion current, are estimated. Using these estimated parameters and all state variables at $T = 1,500$ ms, the response of this neuron was predicted through the end of the data set for this epoch at 6,000 ms. The predicted voltage response for the remainder of Epoch 22 is shown in *blue*. With the exception of the “extra” spike in the predicted voltage response at about 2,600 ms, the prediction matches the data remarkably well

We return to the interesting issue whether ingredients of the full model are identified as not required by the data. Why would we care about this? Our longer-term goal is to use the characterization of individual neurons in a model network circuit for HVC, and if we are able to eliminate degrees of freedom in the model neurons at each node, the overall description of the network model will be simplified. This will make estimations of links in the network and predictions using the network simpler and more efficient.

6.6.2 Neuron 20120517_1_1 Epochs 11 and 12

Next we turn to the stimulating current and resulting voltage response for Epoch 11 of neuron 20110517_1_1; this is in Fig. 6.3. The injected current is again a combination of a chaotic waveform, taken from the Malkus waterwheel equations ([Malkus](#)

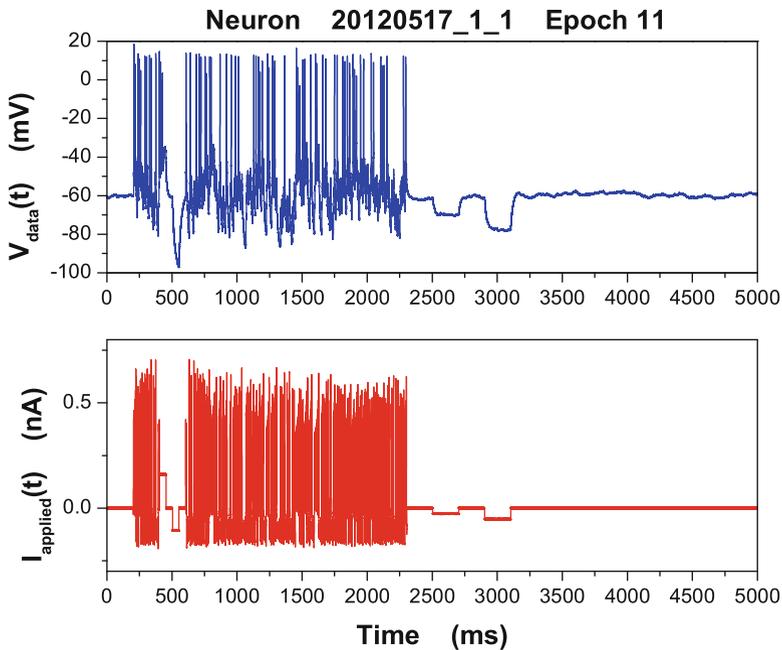


Fig. 6.3 Experimental data for neuron 20120517.1.1 during Epoch 11. Data was collected for five seconds in this epoch but the last 1,500 ms had “zero” injected current. *Top Panel* Membrane voltage response to the injected current shown in the *Bottom Panel*. The injected current is comprised of a complex waveform borrowed from a simple chaotic oscillator with a few sections of constant step currents. The firing rate of this neuron is much higher (Meliza et al. 2013). These two panels are characteristic of the data collected from the many neurons in the experimental efforts. The data for each epoch for each neuron comprises the forces imposed on the neuron by the injected current and the cross membrane voltage observed during an epoch. Estimated errors in the voltage measurements are ± 1 mV or less

1972), and two-step currents embedded in the irregular current near time 500 ms. This neuron spikes much more frequently than the previous neuron for the same approximate amplitude of current presentation, ≈ 0.3 nA. We do not know what class of neuron within HVC we have selected by our procedure of patch clamping neurons with no biophysical selection criterion (Meliza et al. 2013).

We again use $T = 1,500$ ms of the combined $V(t)$ and $I_{\text{applied}}(t)$ data to estimate the 72 parameters and all the gating variables over the time segment $[0$ ms, 1,500 ms]. This estimation for $V(t)$, the only observed state variable, is shown in red in Fig. 6.4. It is followed in the same figure with a prediction for $t > T$ using the estimated parameters in the differential equations along with the values of $V(T)$ and $m(T), h(T)$ for all the currents. This result is shown in blue.

Data from the same neuron is now seen in Fig. 6.5 for the next epoch, Epoch 12. The stimulating current uses a different protocol, and we want to know if the model we have completed through estimating its parameters using data in

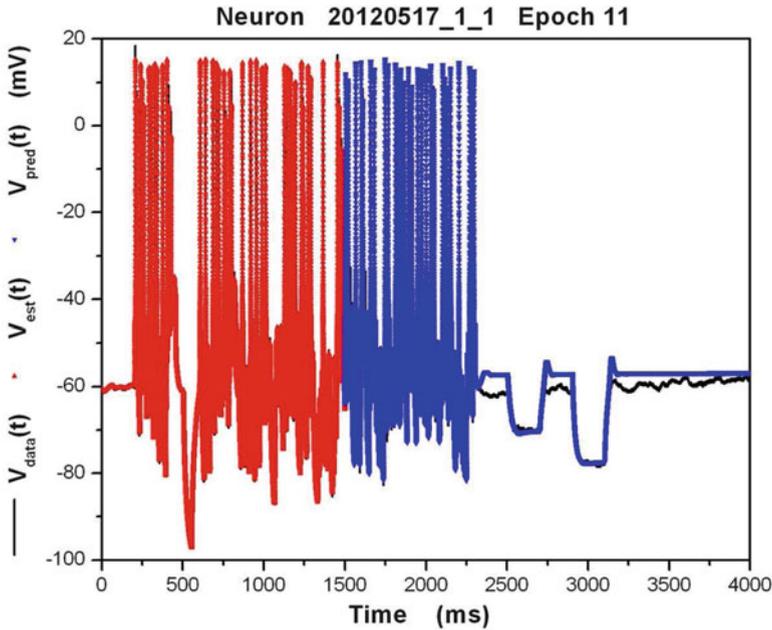


Fig. 6.4 Experimental data, estimated membrane voltage and predicted membrane voltage for neuron 20120517_1_1 during Epoch 11. The observed data is shown as a *heavy black line*. The estimated voltage during the observation window $[0, 1,500]$ ms is shown in *red*. It is during this observation window that the fixed parameters in the model are estimated and the eleven unobserved state variables, here the gating variables $m(t)$ and $h(t)$ for each ion current, are estimated. Using these estimated parameters and all state variables at $T = 1,500$ ms, the response of this neuron was predicted through the end of the data set for this epoch at 6,000 ms. The predicted voltage response for the remainder of Epoch 11 is shown in *blue*

Epoch 11 will predict the voltage response here. To make that prediction, we require as initial conditions $V(t_I)$ and $m(t_I), h(t_I)$ at some time within the epoch. To find $V(t_I), m(t_I), h(t_I)$ we use the model completed by information in the data from Epoch 11. Then fixing the parameters in the model, we determine $V(t_I), m(t_I), h(t_I)$ by the same variational principle used in all these results over an interval $[0 \text{ ms}, 100 \text{ ms}]$ of Epoch 12.

We chose to use 100 ms of the 5,000 ms of data in Epoch 12 to estimate the 12 state variables and then used our completed model from Epoch 11 to predict the remaining 4,900 ms; $t_I = 100$ ms. The result of this is displayed in Fig. 6.6. The estimation of $V(t)$ is in the small segment of red; the prediction is shown in blue. Where the predictions deviate from the observed $V(t)$ for Epoch 12, one is able to see the data in black.

Prediction in another epoch is technically the same kind of test of validity for the completed model within in one epoch assuming that the properties of the neuron are unchanged between the epochs. Over the large data set we collected, we found

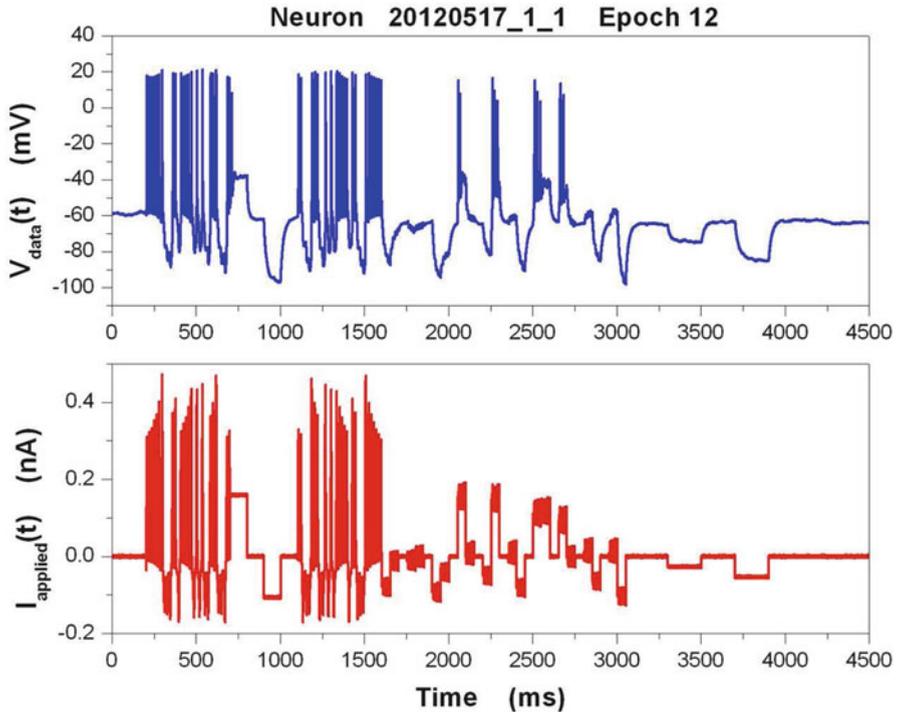


Fig. 6.5 Experimental data for neuron 20120517_1_1 during Epoch 12. Data was collected for five seconds in this epoch but the last 500 ms had “zero” injected current. *Top Panel* Membrane voltage response to the injected current shown in the *Bottom Panel*. The injected current is comprised of a complex waveform borrowed from a simple chaotic oscillator along with several sections of constant step currents. The firing rate of this neuron is much higher than neuron 20110413.4.1 (Meliza et al. 2013). These two panels are characteristic of the data collected from the many neurons in the experimental efforts. The data for each epoch for each neuron comprises the forces imposed on the neuron by the injected current and the cross membrane voltage observed during an epoch. Note that the waveform and the overall amplitude of the injected current differ from that in Epoch 11, Fig. 6.3, while the neuron is the same. Estimated errors in the voltage measurements are ± 1 mV or less

that prediction from models completed in one epoch used to predict in nearby epochs worked much better than epochs that were separated significantly in time. This indicates that the neuron itself was not stationary between those more distant epochs. By using step currents, we are able to estimate the membrane resistance of the neuron from the RC characteristic on presentation of a step, and there is evidence (not shown) that this quantity does change over many epochs.

As enticing as the successful predictions in nearby epochs seem, one should be cautiously optimistic about implementing the network building plan alluded to here of using identified completed neuron models at the nodes of the network. The actual neurons in an experimental network may have different characteristics than

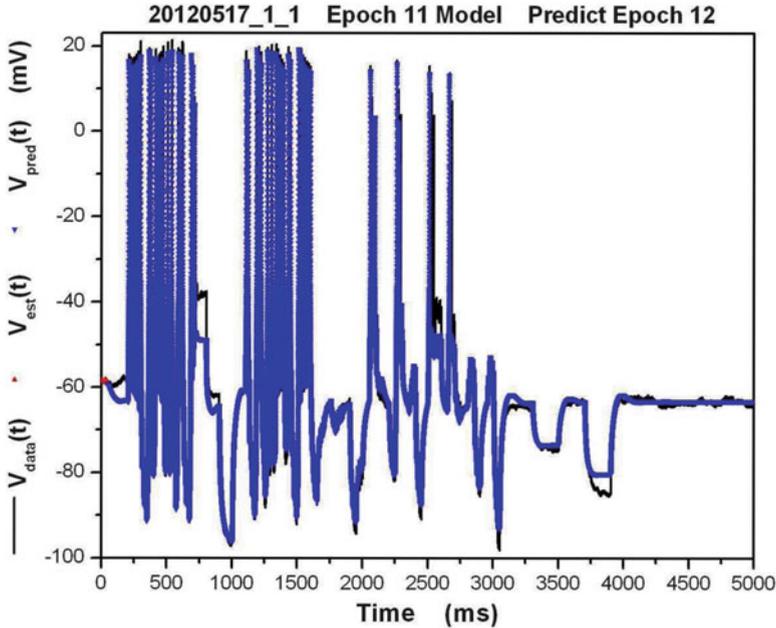


Fig. 6.6 Experimental data, estimated membrane voltage and predicted membrane voltage for neuron 20120517_1_1 during Epoch 12. The observed data is shown as a *heavy black line*. The completed model established in Epoch 11 was used here for estimation and prediction. The estimated voltage during the observation window $[0, 100]$ ms is shown in *red*. (Look carefully, it is a very short segment of data that is used.) It is during this observation window that the voltage $V(t)$ and the gating variables $m(t)$ and $h(t)$ for each ion current are estimated at $T = 100$ ms. Using these estimated parameters and all state variables at $T = 100$ ms, the response of this neuron was predicted through the end of the data set for this epoch at 4,000 ms. The predicted voltage response for the remainder of Epoch 11 is shown in *blue*

the neurons used to complete the model, and the network neurons may have changed because of biophysical events arising from its environment. The plan may yet work, but the present pleasing success does not guarantee it.

It may be that the use of the full path integral to complete neuron models which yields a statistical characterization of the observed neurons will be the appropriate framework for making appropriate statistical predictions on networks.

6.6.3 Neuron 20120406_1_3 Epochs 19 and 15

Without repeating the discussion just completed about estimating and predicting for neuron 20120517_1_1, we display the same kind of result for neuron 20120406_1_3. In this sequence we complete the model in Epoch 19 and then predict both in that

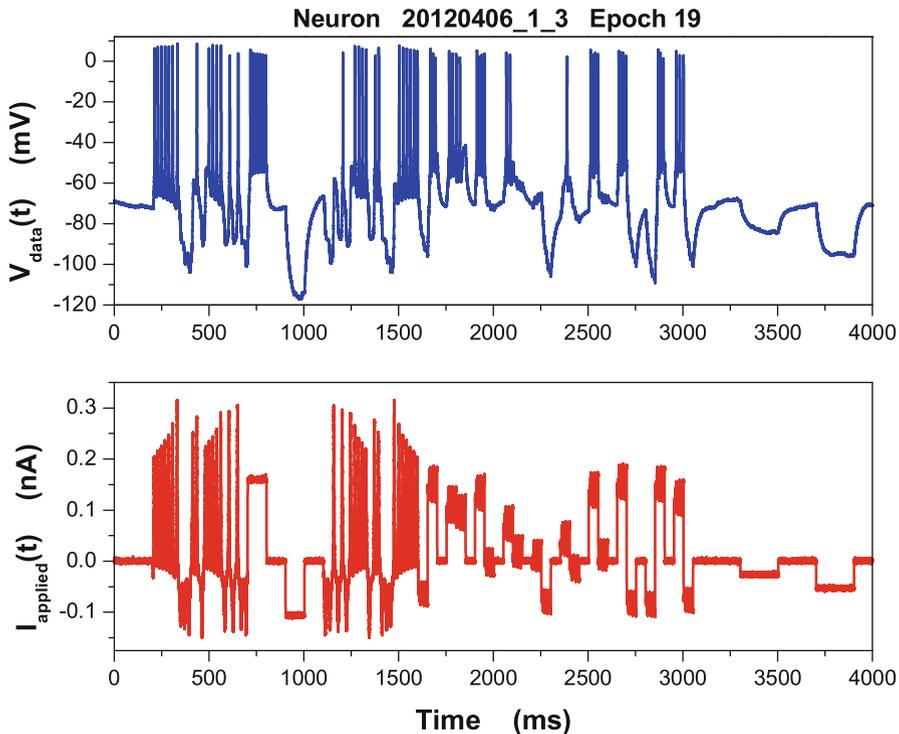


Fig. 6.7 Experimental data for neuron 20120406_1_3 during Epoch 19. Data was collected for four seconds in this epoch. *Top Panel* Membrane voltage response to the injected current shown in the *Bottom Panel*. The injected current is comprised of a complex waveform borrowed from a simple chaotic oscillator along with several sections of nearly constant step currents with some chaotic variation on each step. The firing rate of this neuron is much higher than that of neuron 20110413_4_1 (Meliza et al. 2013). These two panels are characteristic of the data collected from the many neurons in the experimental efforts. The data for each epoch for each neuron comprises the forces imposed on the neuron by the injected current and the cross membrane voltage observed during an epoch. Estimated errors in the voltage measurements are ± 1 mV or less

epoch and in Epoch 15. The data for Epoch 19 are in Fig. 6.7 where we show the injected current and the membrane voltage response. In Fig. 6.8 we exhibit the estimation window and the prediction window with Epoch 19. We display the estimated membrane voltage for the first 1,500 ms of this epoch, and then using the completed model and the estimated value of the state variables at $T = 1,500$ ms, we predict through the rest of the epoch. Figure 6.9 displays the data in Epoch 15 for the same neuron. Figure 6.10 shows the predicted voltage within Epoch 15 using the model completed in Epoch 19. 100 ms of data within Epoch 15 was used to determine the state variables for the Epoch 19 model thus allowing prediction forward within Epoch 15.

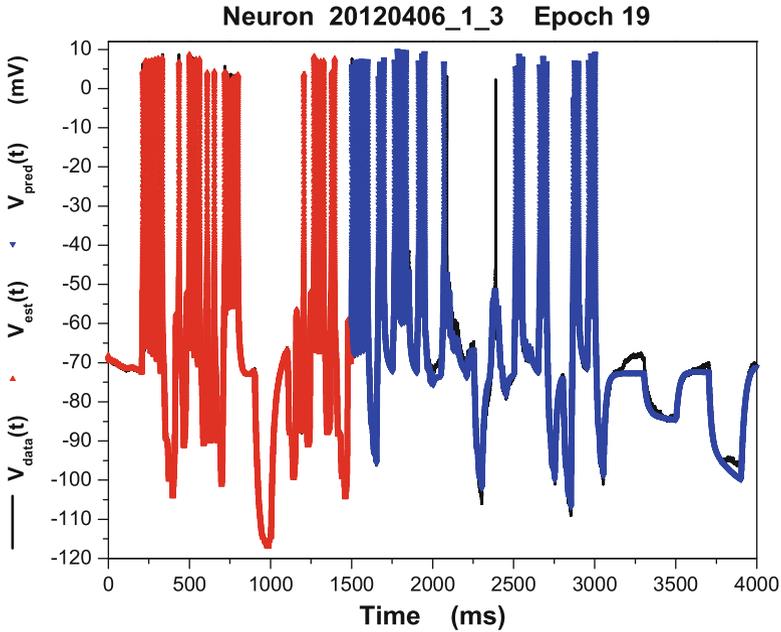


Fig. 6.8 Experimental data, estimated membrane voltage and predicted membrane voltage for neuron 20120406_1_3 during Epoch 19. The observed data is shown as a *heavy black line*. The estimated voltage during the observation window $[0, 1,500]$ ms is shown in *red*. It is during this observation window that the fixed parameters in the model are estimated and the eleven unobserved state variables, here the gating variables $m(t)$ and $h(t)$ for each ion current, are estimated. Using these estimated parameters and all state variables at $T = 1,500$ ms, the response of this neuron was predicted through the end of the data set for this epoch at 4,000 ms. The predicted voltage response for the remainder of Epoch 19 is shown in *blue*

Two comments are of value here:

- (1) The current in Epoch 15 has higher maximum amplitude than in the learning epoch 19. This means that the success in prediction can be associated with a good model and not simply some kind of entrainment of the neuron response.
- (2) In each prediction, within one epoch or across epochs, the success indicates that the full state of the completed model was well estimated by the data assimilation procedure. Prediction examines both the quality of the fixed parameter values and the ability to accurately estimate unobserved state variables.

6.6.4 Estimated Currents and Channel Kinetics

How can we distinguish among the various classes of neuron within HVC using the data we have collected and partially reported on here? One idea is to compare

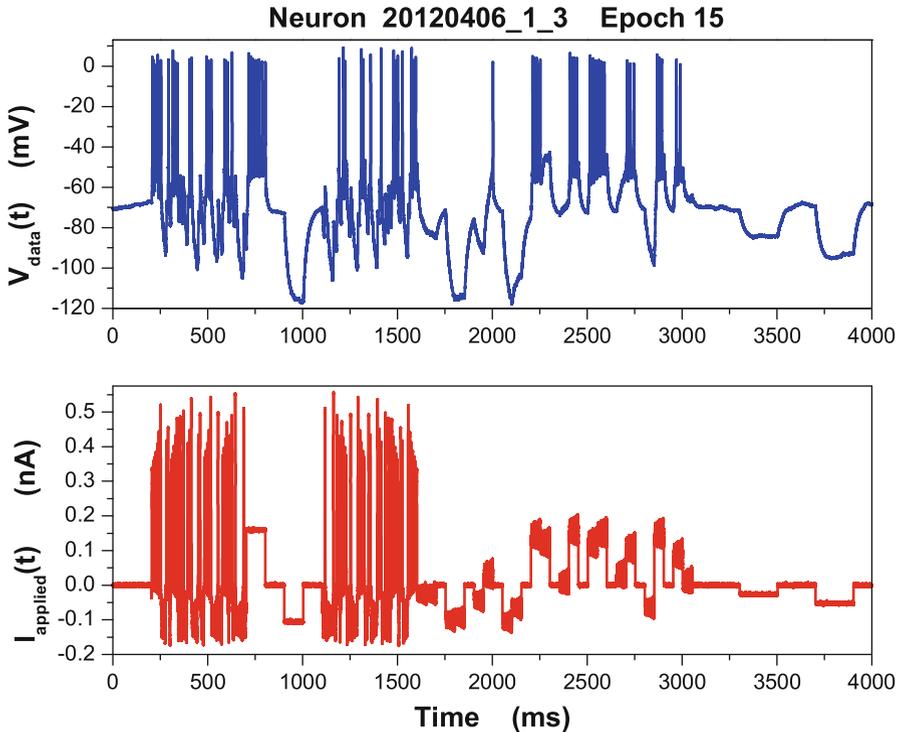


Fig. 6.9 Experimental data for neuron 20120406_1_3 during Epoch 15. Data was collected for four seconds in this epoch. *Top Panel* Membrane voltage response to the injected current shown in the *Bottom Panel*. The injected current is comprised of a complex waveform borrowed from a simple chaotic oscillator along with several sections of nearly constant step currents with some chaotic variation on each step. The firing rate of this neuron is much higher than that of neuron 20110413_4_1 (Meliza et al. 2013). These two panels are characteristic of the data collected from the many neurons in the experimental efforts. The data for each epoch for each neuron comprises the forces imposed on the neuron by the injected current and the cross membrane voltage observed during an epoch. Note that the waveform and the overall amplitude of the injected current differ from that in Epoch 19, while the neuron is the same. Estimated errors in the voltage measurements are ± 1 mV or less

the maximal conductances of each channel. Classifying neurons based upon their collection of channel maximal conductances may not accurately reflect the dynamical importance of the channels on driving the membrane voltage over a significant time interval. It is not the maximal conductance alone but the strength of the ion current that influences changes in the membrane voltage. Large driving currents may result from small conductances for a channel which has a very slow time constant.

The ion currents, not just one or another constant parameters comprising those currents, are the forces driving the neuron. This suggests that the sensitivity of optimization procedures such as we used in the analysis of these data to particular values of certain parameters may miss the core physical question about the forces

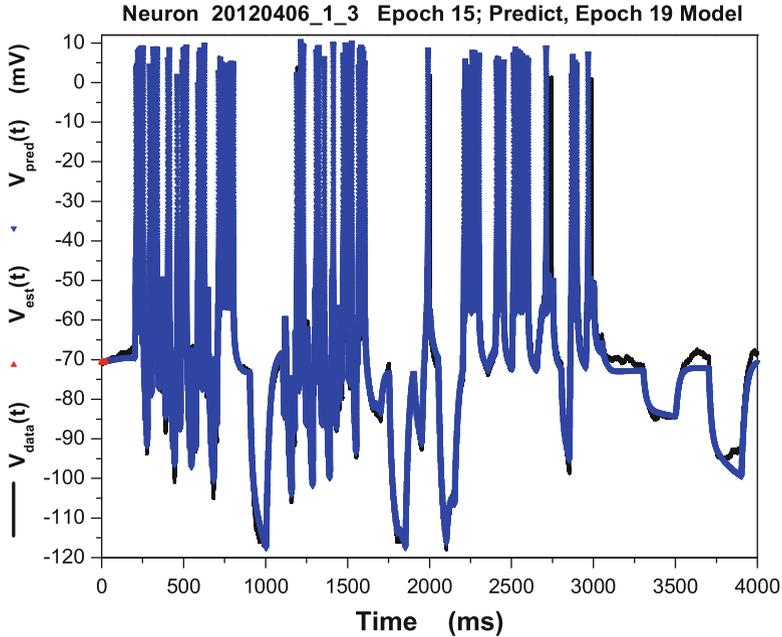


Fig. 6.10 Experimental data, estimated membrane voltage and predicted membrane voltage for neuron 20120406_1_3 during Epoch 15. The observed data is shown as a *heavy black line*. The completed model established in Epoch 19 was used here for estimation and prediction. The estimated voltage during the observation window $[0, 100]$ ms is shown in *red*. (Look carefully, it is a very short segment of data that is used.) It is during this observation window that the voltage $V(t)$ and the gating variables $m(t)$ and $h(t)$ for each ion current are estimated at $T = 100$ ms. Using these estimated parameters and all state variables at $T = 100$ ms, the response of this neuron was predicted through the end of the data set for this epoch at 4,000 ms. The predicted voltage response for the remainder of Epoch 15 is shown in *blue*

acting on the neuron. The matter of sensitivity to time-dependent forces of numerical procedures is the issue. One can recall that in the consideration of tidal forces on the Earth's ocean, even though Jupiter's mass is much larger than the moon's mass, it is the larger force of the moon, because of the shorter distance, that is the determining physical ingredient.

None of the models indicated a strong contribution from the I_h current, nor was there a strong contribution from L-type Ca channels.

As we will use the biophysical characterizations of these individual neurons to construct networks of these neurons in further analyses of the birdsong system, surveying the important currents will allow us to reduce the size of the model neurons used at these nodes and streamline the computations required for predicting network behavior. This is a dynamical model reduction method (Meliza et al. 2013).

6.7 Comments on the Analysis of These Data

Beyond the ability of the statistical data assimilation formulation to assist in the development and completion of predictive, quantitative biophysical models, there is substantial information in the actual parameters for the observed neurons and in the specific ion currents captured within models for each neuron (Meliza et al. 2013).

Two broad questions about the results of the estimates arise:

- For a given neuron, multiple parameter sets might be able to produce models with similar high-quality predictions. This is in part because the real, physical neuron has many more degrees of freedom than the proposed model, and therefore, a single parameter set from this or another approximation to the data assimilation path integral simply may not exist. The projection from the “actual” high-dimensional state space of the real neuron may lead to an apparent statistical result of this kind.
- A second source of variability is that neurons may be able to achieve the same physiological behavior through many different combinations and expression patterns of channels (Prinz et al. 2004). Given the indeterminism of the models, it is difficult to know to what extent neurons take advantage of this flexibility, though studies of single-neuron gene expression suggest this may be the case (Schulz et al. 2006). Further complicating this question is the existence of compensatory mechanisms that allow at least some types of neurons to maintain specific behaviors even when specific channels are absent (Swensen and Bean 2005).

6.8 Synopsis and Perspectives: Analysis of Experimental Data

The particular example discussed in this chapter is likely to be representative of the use of the methods developed in this book when one comes to examine laboratory and field data for complex systems.

The use of twin experiments in identifying how many and which measurements are required for any proposed model for the source of the observed data should be a valuable tool over many different classes of problem. While no laboratory or field data is required to make these estimates, as more and more complicated models are explored for complex system behavior, these estimates will be a good guide to how the measurement needs will play a role in establishing quantitative aspects of models. Unfortunately, at this stage, if one finds that the number of required measurements is insufficient, no remedies are presently available, except more measurements, of course.

If one is in the advantageous position of having sufficient observations taken often enough in time, then the statistical data assimilation procedure as exemplified in this chapter provides a tool for establishing consistency between the proposed

model and the available data. As in the discussion presented, details matter, and one should focus on those details both about the experiments and field observations as well as about the models themselves. The first are certain to be noisy and, even when sufficient, sparse. The latter will have uncertainties that must be wished away—no model errors, thank you—or quantified in some manner, perhaps by a consistency argument.

Of course, it is possible to propose models and come away impressed with a “fit” to selected data, but coming to grips with the details and predicting beyond existing measurements is eventually the key one wishes to grasp. Attention to the biophysics or geophysics or chemical physics of the processes at work in producing the data is certain to be ever important.

Chapter 7

Unfinished Business

Prologue to the Chapter

There are many more questions raised for future investigation than are answered in a monograph such as this. This final chapter is meant to suggest what is yet to be addressed.

Perhaps the core result in this book is the exact formulation for evaluating the conditional probability distribution $P(\mathbf{x}(m)|\mathbf{Y}(m))$ of the D -dimensional state $\mathbf{x}(m)$ of a model of a physical system at the end of an observation (or training or assimilation) time interval $[t_0, t_m = T]$ when L -dimensional measurements $y_l(t_n)$; $l = 1, 2, \dots, L$; $n = 0, 1, \dots, m$ are made at times $\{t_0, t_1, \dots, t_m\}$. The measurements until time t_n are collected into $\mathbf{Y}(n) = \{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(n)\}$. From $P(\mathbf{x}(m)|\mathbf{Y}(m))$ we arrive at an expression for the expected value of any function $G(\mathbf{X})$ along the **path** $\mathbf{X} = \{\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_m)\} = \{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(m)\}$ traversed by the model

$$E[G(\mathbf{X})|\mathbf{Y}(m)] = \frac{\int d\mathbf{X} G(\mathbf{X}) \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]}{\int d\mathbf{X} \exp[-A_0(\mathbf{X}, \mathbf{Y}(m))]} \quad (7.1)$$

in which

$$\begin{aligned} -A_0(\mathbf{X}, \mathbf{Y}(m)) &= \sum_{n=0}^m \log[P(\mathbf{y}(n)|\mathbf{x}(n), \mathbf{Y}(n-1))] \\ &+ \sum_{n=0}^{m-1} \log[P(\mathbf{x}(n+1)|\mathbf{x}(n))] + \log[P(\mathbf{x}(0))]. \end{aligned} \quad (7.2)$$

By well-selected choices of $G(\mathbf{X})$ one can determine the mean trajectory, variations about that mean, marginal distributions of states $x_a(t_n)$; $a = 1, 2, \dots, D$, and

whatever other information one wishes about the distribution of state variables conditioned on measurements.

This formula places consideration of paths in state space at the center of ideas about assimilating information from data into models of simple or complex systems. It also focuses attention on the need to evaluate high-dimensional, $(m + 1)D$ -dimensional, integrals as the essential calculational problem. The necessity of doing this integral does not inform us how to make a physical model of observed phenomena nor does it inform us what to measure and how often.

In considering performing this integral by saddle path methods (4DVar and its variations in the geophysical literature) or direct Monte Carlo methods of one sort or another one must still address at least the following questions:

1. How many measurements at each observation time t_n are required to allow accurate estimation of the integral?
2. How often in time must those measurements be performed?
3. What are we to do in the absence of enough measurements?
4. What level of noise in the measurements allows accurate estimation of the path integral?
5. What is the criterion for the quality of a completed model in which all fixed parameters are now estimated? We have argued that quality of prediction using the model under changed conditions of forcing is the metric to use. What precise mathematical criterion should one adopt for that?
6. How does one quantify model errors in the specification of the transition probability $P(\mathbf{x}(n + 1)|\mathbf{x}(n))$, especially when the error is the absence of a term in the dynamical equations? What tests for consistency of one or another assumption about model errors are relevant?
7. On what machines and with what massively parallel computing methods should we take on compelling questions about big complex systems—weather prediction, gene regulation, nervous system characterization, reactive fluid flow in complex enclosures, ...?

This list, perhaps augmented by additions from the reader, could be the conclusion of this overall discussion. However on the third item we have a suggestion, on the first two items some comments.

On the first item, we note two ideas in this regard:

- When there are no model errors one can establish, how many measurements L_c are required of a model using twin experiments to determine the value of L through evaluating when two signals from the model with different initial conditions synchronize? Within this question is also the need to determine which measurements as well as how many.
- Within the path integral when model errors are present, how many measurements are required to make the action $A_0(\mathbf{X}, \mathbf{Y}(m))$ have a single minimum?

We have given examples of this, but what is required is an algorithmic production code to evaluate these items given a proposed model.

On the second item, we suggest it may be enough to estimate the largest Lyapunov exponent of the model incorporated into the transition probability $P(\mathbf{x}(n+1)|\mathbf{x}(n))$. This will be a quantity which can differ in various regions of state space \mathbf{x} space. It may then be sufficient to assure that the largest value of $t_{n+1} - t_n$ does not exceed the inverse of this largest Lyapunov exponent.

On the third item, we turn briefly to an idea from nonlinear dynamics.

7.1 “More Measurements”: The Use of Time Delay Phase Space Reconstruction

The issue of adding measurements when the number of observations is not enough to assure stability on the synchronization manifold or, equivalently, the action does not have a clear stable minimum has been only touched on in the main part of this text. Clearly one option is to make more measurements. Another option, which we shall address here, is to use more information about the waveform of the time series of the measurements. The ideas we discuss in this section come from considerations of “embedding” that are quite common in the nonlinear dynamics literature; naturally, they have their own issues.

The model dynamical system works in a D -dimensional state space with state vectors at each time t_n : $\mathbf{x}(t_n) = \mathbf{x}(n)$. The deterministic dynamics for the model takes states $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$. The dynamics is Markov as only $\mathbf{x}(n)$ is required to know $\mathbf{x}(n+1)$ through our rule. \mathbf{p} is a set of parameters. The dynamical rule $\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{p})$ is the discrete time version of a D -dimensional set of ordinary differential equations that could have come from PDEs. The rule may well be the implementation of an algorithm for solving the ODEs.

As we observe L state variable-related quantities at each t_n , $D - L$ state variables remain unobserved, and our goal is to use the information in the measurements $\mathbf{Y}(m) = \{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(m)\}$ to estimate the $D - L$ unobserved state variables and the NP fixed parameters \mathbf{p} . These should be estimated throughout the observation window and especially at the end $t_m = T$. If we have estimates of $\mathbf{x}(T)$ and \mathbf{p} , we may use the rule to **predict** for $t > T$.

The situation we want to look at in this discussion arises when we have made as many measurements as we are able, and this number is not enough to stabilize the synchronization manifold, should we be considering deterministic (no model error) dynamics or, equivalently, the number of observations to assure that the action $A_0(\mathbf{X})$ has a clear, isolated minimum. We need more information from the observations we are able to extract from temporally isolated single measurements.

In the study of nonlinear systems it has long been recognized (Sauer et al. 1991; Abarbanel 1996; Kantz and Schreiber 2004) that sampling the waveform of a time series $s(t_n) = s(n)$ and recording not only $s(n)$ but also $s(n - 1)$ and other time-

delayed values of $s(n)$ along the observed waveform provides useful independent looks at the information in the measurement. What follows now arose in discussions with Ulrich Parlitz.

While this is discussed at some length in the references, the basic idea is that we characterize independent degrees of freedom in dynamics with derivatives in time which could be higher and higher order. Since we cannot measure in continuous time, we are restricted to approximating time derivatives by finite differences, and for example, if we know $s(t_n)$ and wish to approximate $\frac{ds(t)}{dt}$, we might start with $\frac{ds(t)}{dt} \approx \frac{s(t_{n+1}) - s(t_n)}{t_{n+1} - t_n}$, and the only new information in this approximation, relative to what we know at t_n , namely, $s(t_n)$, is $s(n+1) = s(t_{n+1})$, so why not use $s(n+1)$ as a coordinate in place of ‘ $\frac{ds(t)}{dt}$ ’? In the examination of this basic idea, there follows a discussion of what time delay to use and how many such time delays produce an equivalent space to the original coordinates. We skip this discussion, accepting the conclusions (Abarbanel 1996; Kantz and Schreiber 2004; Sauer et al. 1991).

Utilizing all of that, we have in mind that each of the L observations will be of some scalar function of the state variables $h(\mathbf{x})$ evaluated at time t_n and at $K - 1$ earlier times, namely, times $t_n - (k - 1)\Delta t$; $k = 1, 2, \dots, K$. So when we make the observations $y_l(n)$ we have in mind to associate them with the values $\{h(\mathbf{x}(n)), h(\mathbf{x}(n - 1)), h(\mathbf{x}(n - 2)), \dots, h(\mathbf{x}(n - (K - 1)))\}$. Here Δt is taken as a unit of time. Given the observations we want to “tune up” the model so that it produces values $\mathbf{x}(n)$ at a time t_n that yields $h(\mathbf{x}(n)) \approx y_1(n)$, $h(\mathbf{x}(n - 1)) \approx y_2(n)$, and so forth until $h(\mathbf{x}(n - (K - 1))) \approx y_K(n)$.

The discussion now proceeds with $L = 1$, namely, a situation where only one measurement is made at each time. This is for simplicity of notation. If one makes L measurements at each t_n , the formulation here can be extended to each of the L observed time series corresponding to the measurement functions $h_l(\mathbf{x}(n))$; $l = 1, 2, \dots, L$.

The idea is now that this problem can be formulated in a convenient manner when we use coordinates made out of the measurement function $h(\mathbf{x})$ by creating a new space “equivalent” in the usual sort of way (Abarbanel 1996; Kantz and Schreiber 2004) to the original \mathbf{x} space. We call vectors in that space \mathbf{S} and they have K components.

The vectors \mathbf{S} are derived from the vectors \mathbf{x} through a transformation $S_k = \phi_k(\mathbf{x})$; $k = 1, 2, \dots, K$, and the $\phi_k(\mathbf{x})$ are found from the definition $\mathbf{S}(n) = \{s_1(n), s_2(n), \dots, s_K(n)\} = \{h(\mathbf{x}(n)), h(\mathbf{x}(n - 1)), h(\mathbf{x}(n - 2)), \dots, h(\mathbf{x}(n - (K - 1)))\} = \phi(\mathbf{x})$. One consequence of the discussion about time delay coordinates that we note is that if $K \geq 2D + 1$, that is a sufficient, but not a necessary, condition for the equivalence of the \mathbf{S} and the \mathbf{x} spaces.

Now we want to take the rule $\mathbf{x}(n) \rightarrow \mathbf{x}(n + 1) = \mathbf{f}(\mathbf{x}(n))$ and create a rule for $\mathbf{S}(n) \rightarrow \mathbf{S}(n + 1)$. The $\mathbf{S}(n)$ dynamics is where we will do the state and parameter estimation for the unobserved \mathbf{x} and the \mathbf{p} . (We don’t always show \mathbf{p} now.)

This dynamical rule for $\mathbf{S}(n + 1)$ can be read from

$$\begin{aligned}
s_1(n+1) &= h(\mathbf{x}(n+1)) = h(\mathbf{f}(\mathbf{x}(n))) \\
s_2(n+1) &= h(\mathbf{x}(n-1+1)) = h(\mathbf{x}(n)) = s_1(n) \\
s_3(n+1) &= h(\mathbf{x}(n-2+1)) = h(\mathbf{x}(n-1)) = s_2(n) \\
&\vdots \\
s_k(n+1) &= s_{k-1}(n) \\
&\vdots \\
s_K(n+1) &= s_{K-1}(n).
\end{aligned} \tag{7.3}$$

All this is good, but we have yet to deal with eliminating $\mathbf{x}(n)$ in the expression for $s_1(n+1)$. We have defined the “forward” change of variables $\mathbf{x} \rightarrow \mathbf{S} = \phi(\mathbf{x})$, and now we want the reverse transformation $\mathbf{S} \rightarrow \mathbf{x} = \psi(\mathbf{S})$. How to find this and even to establish if it exists needs to be thought about. If, however, all is well, then we may write

$$s_1(n+1) = h(\mathbf{f}(\mathbf{x}(n))) = h(\mathbf{f}(\psi(\mathbf{S}(n)))) = \chi(\mathbf{S}(n)), \tag{7.4}$$

and we are now in possession of a dynamical rule in our new space, namely: $\mathbf{S}(n) \rightarrow \mathbf{S}(n+1) = \{\chi(\mathbf{S}(n)), s_1(n), s_2(n), \dots, s_{K-1}(n)\} = \mathbf{g}(\mathbf{S}(n))$.

Our original problem in \mathbf{x} space has now become the problem in \mathbf{S} space: we observe L time series over the observation window $[0, T]$ and want to determine the $K - L$ unobserved states over this window and all fixed parameters in the rule $\mathbf{S}(n) \rightarrow \mathbf{S}(n+1)$. If there are fixed parameters in the measurement function $h(\mathbf{x})$, we throw them into the pot as well.

If we accomplish this and can then predict $\mathbf{S}(t)$ for $t > T$, then we can extract the answer to our original question in \mathbf{x} space through the use of $\mathbf{x}(n) = \psi(\mathbf{S}(n))$ and the knowledge (estimation) we now have of the \mathbf{p} .

A long, interesting discussion of the inverse transformation $\mathbf{S} \rightarrow \mathbf{x}$ could now ensue, and it should include ideas, perhaps some practical ideas, on how we will implement $\mathbf{x} = \psi(\mathbf{S})$ in a stable numerical scheme.

The estimation procedure will ask that the output of the model dynamics in \mathbf{S} space be such that the L measurements $y_l(n)$ at each measurement time t_n should approximate the values $\{s_1(n), s_2(n), \dots, s_L(n)\}$, and this can be done point by point or by minimizing something like a metric (cost function)

$$\sum_{n=0}^m \left\{ \sum_{l=1}^L (y_l(n) - s_l(n))^2 \right\}, \tag{7.5}$$

subject to $\mathbf{S}(n+1) = \text{known map}(\mathbf{S}(n))$ from above.

Absent here is any discussion of noise in the measurements and errors in the models. From all of the earlier presentation, however, we can formulate the required path integral and its approximations in \mathbf{S} space with no problem.

What did we gain with all this? More measurements with information about the temporal development of the source of data without making more observations! The key notion is that in the waveform of what we observe, there is much more information than just the instantaneous values $s(n) = h(\mathbf{x}(n))$. The actual dynamics takes place in a larger space than that of the $L < D$ observations and is projected down to the L -dimensional space in which we record data at each measurement time t_n . Embedding using time delays or other means reverses that projection geometrically and unveils dynamical information ([Abarbanel 1996](#); [Kantz and Schreiber 2004](#)).

Bibliography

- Abarbanel, H.D.I.: Analysis of Observed Chaotic Data. Springer, New York (1996)
- Abarbanel, H.D.I.: Effective actions for statistical data assimilation. Phys. Lett. A **373**, 4044–4048 (2009). doi: 10.1016/j.physleta.2009.08.072
- Abarbanel, H.D.I., Creveling, D., Gill, P.E.: State and parameter estimation in nonlinear systems as an optimal tracking problem. Phys. Lett. A **372**, 2640–2644 (2008)
- Abarbanel, H.D.I., Creveling, D.R., Farsian, R., Kostuk, M.: Dynamical State and parameter estimation. SIAM J. Appl. Dynam. Syst. **8**, 1341–1381 (2009)
- Abarbanel, H.D.I., Kostuk, M., Whartenby, W.: Data assimilation with regularized nonlinear instabilities. Quart. J. R. Meteorol. Soc. **136**, 769–783 (2010)
- Abarbanel, H.D.I., Bryant, P., Gill, P.E., Kostuk, M., Rofeh, J., Singer, Z., Toth, B., Wong, E.: Dynamical parameter and state estimation in neuron models, Chapter 8. In: Ding, M., Glanzman, D.L. (eds.) The Dynamic Brain, pp. 139–180. Oxford University Press, Oxford (2011)
- Alexander, F.J., Eyink, G.L., Restrepo, J.M.: Accelerated Monte Carlo for optimal estimation of time series. J. Stat. Phys. **119**, 1331–1345 (2005)
- Apte, A., Hairer, M., Stuart, A., Voss, J.: Sampling the posterior: An approach to non-Gaussian data assimilation. Phys. D: Nonlinear Phenom. **230**, 50–64 (2007). doi: 10.1016/j.physd.2006.06.009
- Auletta, G., Fortunato, M., Parisi, G.: Quantum Mechanics. Cambridge University Press, Cambridge (2009). ISBN 978-0-521-86963-8
- Barclay, A., Gill, P.E., Rosen, J.B.: SQP methods in optimal control. In: Bulirsch, R., Bittner, L., Schmidt, W.H., Heier, K. (eds.) Variational Calculus, Optimal Control and Applications. International Series of Numerical Mathematics, vol. 124, pp. 207–222. Birkhaeuser, Basel (1998)
- Batchelor, G.K.: An Introduction to Fluid Dynamics. Cambridge University Press, Cambridge (1967)
- Bazley, N.W., Fox, D.W.: Error bounds for expectation values. Rev. Mod. Phys. **35**, 712–715 (1963)
- Beard, D.A.: A biophysical model of the mitochondrial respiratory system and oxidative phosphorylation. PLoS Comput. Biol. **1**, 252–264 (2005). doi: 10.1371/journal.pcbi.0010036
- Burrage, P.M.: Numerical methods for stochastic differential equations. Ph.D. thesis, University of Queensland (1999)
- Burrage, P.M., Burrage, K.: A variable stepsize implementation for stochastic differential equations. SIAM J. Sci. Comput. **24**, 848–864 (2002)
- Carroll, T., Pecora, L.: Synchronization in chaotic systems. Phys. Rev. Lett. **64**, 821–824 (1990)
- Caves, C.M.: Quantum mechanics of measurements distributed in time: A path-integral formulation. Phys. Rev. D **33**, 1643–1665 (1986)

- Chandler, D.: *Introduction to Modern Statistical Mechanics*, p. 175. Oxford University Press, Oxford (1987)
- Chapman, S.: On the Brownian displacements and thermal diffusion of grains suspended in a non-uniform fluid. *Proc. Roy. Soc. Ser. A* **119**, 34–54 (1928)
- Colpitts, E.H.: Oscillation generator. US Patent 1,624,537, filed: 1 Feb 1918; issued: 12 April 1927
- Cox, H.: On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Automat. Contr.* **AC-9**, 5–12 (1964)
- Curtis, F.E., Schenk, O., Wächter, A.: An interior-point algorithm for large scale nonlinear optimization with inexact step computations. *SIAM J. Sci. Comput.* **32**, 3447–3475 (2010)
- Debye, P.: Näherungsformeln für die zylinderfunktionen für grosse werte des arguments und unbeschränkt veränderliche werte des index. *Math. Ann.* **67**(4), 535–558 (1909). doi:10.1007/BF01450097 (English translation in Debye, P.J.W.: *The collected papers of P.J.W. Debye*. Interscience Publishers Inc., New York (1954). ISBN 978-0-918024-58-9, MR0063975)
- De Feo, O., Maggio, G.M.: Bifurcations in the Colpitts oscillator: From theory to practice. *Int. J. Bifurcat. Chaos.* **13**, 2917–34 (2003)
- Dochain, D.: State and parameter estimation in chemical and biochemical processes: a tutorial. *J. Process Contr.* **13**, 801–818 (2003). doi: 10.1016/S0959-1524(03)00026-X
- Ebers, J.J., Moll, J.L.: Large signal behaviour of junction transistors. *Proc. I. R. E.* **42**, 1761–1772 (1954)
- Elson, R., Huerta, R., Rulkov, N.F., Abarbanel, H.D.I., Rabinovich, M.I.: Synchronous behavior of two coupled biological neurons. *Phys. Rev. Lett.* **81**, 5692–5695 (1998)
- Evensen, G.: *Data Assimilation: The Ensemble Kalman Filter*, 2nd edn. Springer, Berlin (2009). ISBN: 978-3-642-03710-8
- Fano, R.M.: *Transmission of Information: A Statistical Theory of Communication*. Wiley and M.I.T. Press, Cambridge (1961)
- Feller, W.: *An Introduction to Probability Theory and Its Applications*. Wiley, New York (1971). ISBN 10: 0471257095 / 0-471-25709-5; ISBN 13: 9780471257097
- Feynman, R.P., Hibbs, A.R.: *Quantum Mechanics and Path Integrals*. McGraw-Hill Companies Inc., New York (1965)
- Feynman, R.P., Vernon, F.L.: The theory of a general quantum system interacting with a linear dissipative system. *Ann. Phys.* **24**, 118–173 (1963)
- Friedland, B., Bernstein, I.: Estimation of the state of a nonlinear process in the presence of nongaussian noise and disturbances. *J. Franklin Inst.* **281**, 455–480 (1966)
- Faragó, I., Georgiev, K., Havasi, Á. (eds.): *Proceedings of the NATO Advanced Research Workshop on Advances in Modeling for Environmental Security*, Borovetz, Bulgaria, 8–12 May 2004. *Earth and Environmental Sciences*, vol. 54. Springer, Dordrecht (2005). ISBN 10 1-4020-3349-4
- Gamerman, D., Lopes, H.F.: *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Texts in Statistical Science. Chapman and Hall, Boca Raton (2006)
- Gill, A.E.: *Atmosphere-Ocean Dynamics*. International Geophysics Series, vol. 30. Academic, Orlando (1982)
- Gill, P.E., Wright, M.H.: *Practical Optimization*. Academic, London (1982). ISBN-10: 0122839528, ISBN-13: 9780122839528
- Gill, P.E., Murray, W., Saunders, M.A.: User's Guide for SNOPT 5.3: A Fortran Package for Large-Scale Nonlinear Programming, p. 67. Stanford Optimization Laboratory, Stanford (1998)
- Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM J. Optim.* **12**, 979–1006 (2002)
- Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.* **47**, 99–131 (2005)
- Graham, L.: Modelling neuronal biophysics. In: Arbib, M. (ed.) *The Handbook for Brain Theory and Neural Networks*, pp. 164–170, 2nd edn. MIT Press, Cambridge (2002)
- Gummel, H.K., Poon, R.C.: An integral charge control model of bipolar transistors. *Bell Syst. Tech. J.* **49**, 827–852, (1970)

- Hamill, T.M.: Ensemble-based atmospheric data assimilation: A tutorial. In: Palmer, T., Hagedorn, R. (eds.) *Predictability of Weather and Climate*, pp. 124–156. Cambridge University Press, Cambridge (2006)
- Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J.: Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv Euro. J. Physiol.* **391**, 85–100 (1981). doi: 10.1007/BF00656997
- Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970). <http://www.jstor.org/stable/2334940>
- Hille, B.: *Ionic Channels of Excitable Membranes*, 2nd edn. Sinauer Associates, Sunderland (1992)
- Hochberg, D., Molina-Paris, C., Pérez-Mercader, J., Visser, M.: Effective action for stochastic partial differential equations. *Phys. Rev. E* **60**, 6343–6360 (1999)
- Horváth, A., Manini, D.: Parameter estimation of kinetic rates in stochastic reaction networks by the EM method. In: *International Conference on BioMedical Engineering and Informatics* (2008). doi.ieeecomputersociety.org/10.1109/BMEI.2008.237
- Huang, D.: Synchronization-based estimation of all parameters of chaotic systems from time series. *Phys. Rev. E* **69** 067201 (2004)
- Huguenard, J.R., McCormick, D.A.: Simulation of the currents involved in rhythmic oscillations in thalamic relay neurons. *J. Neurophysiol.* **68**, 1373–83 (1992)
- Huys, Q.J., Paninski, L.: Smoothing of, and parameter estimation from, noisy biophysical recordings. *PLoS Comput. Biol.* **5**, e1000379 (2009). doi: 10.1371/journal.pcbi.1000379
- Huys, Q.J., Ahrens, M.B., Paninski, L.: Efficient estimation of detailed single-neuron models. *J. Neurophysiol.* **96**, 872–90 (2006)
- Illing, L., Fordyce, R., Saunders, A., Ormond, B.: Experiments with a Malkus-Lorenz water wheel: chaos and synchronization. *Am. J. Phys.* **80**, 192–202 (2012). doi: 10.1119/1.3680533
- Illing, L., Saunders, A.M., Hahs, D.: Multi-parameter identification from scalar time series generated by a Malkus-Lorenz water wheel. *CHAOS* **22**, 013127 (2012). <http://dx.doi.org/10.1063/1.3689441>
- Jazwinski, A.H.: *Stochastic Processes and Filtering Theory*, p. 376. Academic, New York (1970); Dover Publications, New York (2007)
- Johnston, D., Wu, S.M.S.: *Foundations of Cellular Neurophysiology*. MIT Press, Cambridge (1995). ISBN: 0 262 10053 3
- Jouvet, B., Pythian, R.: Quantum aspects of classical and statistical fields. *Phys. Rev. A* **19**, 1350–1355 (1979)
- Kalnay, E.: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge (2003)
- Kantz, H., Schreiber, T.: *Nonlinear Time Series Analysis*, 2nd edn. Cambridge University Press, Cambridge (2004)
- Kennedy, M.P.: Chaos in the Colpitts oscillator. *IEEE Trans. Circuit. Syst. I: Fund. Theor. Appl.* **41** 771–774 (1994)
- Kirk, D.E.: *Optimal Control Theory: An Introduction*. Dover Publications Inc., Mineola (1998) (original published by Prentice-Hall in 1970)
- Klauder, J.R., Petersen, W.P.: Spectrum of certain non-self-adjoint operators and solutions of Langevin equations with complex drift. *J. Stat. Phys.* **39**, 53–72 (1985)
- Kloeden, P.E., Platen, E.: *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin (1992)
- Koch, C.: *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, Oxford (1999)
- Kolmogorov, A.N.: Ueber die analytischen methoden in der wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 415–458 (1931)
- Konnur, R.: Synchronization-based approach for estimating all model parameters of chaotic systems. *Phys. Rev. E* **67** 027204 (2003)
- Kostuk, M.: Ph.D. dissertation, University of California, San Diego (2012)
- Laje, R., Mindlin, G.B.: *The Physics of Birdsong*. Oxford University Press, Oxford (1999); Springer, New York (2005)

- Laplace, P.S.: Memoir on the probability of causes of events. *Mém. Math. Phys.* **16**, (1774) [English translation by Stigler, S.M.: *Stat. Sci.* **1**, 364–378 (1986)]
- Laurent, G., Stopfer, M., Friedrich, R.W., Rabinovich, M.I., Volkovskii, A., Abarbanel, H.D.I.: Odor encoding as an active dynamical process: experiments, computation, and theory. *Annu. Rev. Neurosci.* **24**, 263–297 (2001)
- Lindenberg, K., West, B.J.: *The Nonequilibrium Statistical Mechanics of Open and Closed Systems*. Wiley, VCH Publishers, New York (1990). ISBN-10: 0-471-18683-X; ISBN-13: 978-0-471-18683-0
- Lindley, D.V.: On a measure of the information provided by an experiment. *Ann. Math. Stat.* **27**, 986–1005 (1956)
- Lorenç, A.C., Payne, T.: 4D-Var and the butterfly effect: Statistical four-dimensional data assimilation for a wide range of scales. *Q. J. R. Meteorol. Soc.* **133**, 607–14 (2007) (There are numerous references to 4DVar, and we selected this one as it critically examines the method after a pedagogical introduction to it)
- Lorenz, E.N.: Deterministic, nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963)
- Lorenz, E.N.: Predictability: A problem partly solved. In: *Proceedings of the Seminar on Predictability*, vol. 1, pp. 1–18. ECMWF, Reading (1996)
- Lorenz, E.N., Emanuel, K.A.: Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414 (1998)
- Luenberger, D.G.: Observing the state of a linear system. *IEEE Trans. Mil. Electron.* **3MIL**, 74–80 (1964)
- Luenberger, D.G.: Observers for multivariable systems. *IEEE Trans. Automat. Contr.* **AC11**, 190–197 (1966)
- Luenberger, D.G.: An introduction to observers. *IEEE Trans. Automat. Contr.* **AC-16**, 396–602 (1971)
- Luenberger, D.G.: *Introduction to Dynamic Systems, Theory, Models and Applications*. Wiley, New York (1979)
- Lyons, M.A., Yang, R.S.H., Mayeno, A.N., Reisfeld, B.: Computational toxicology of chloroform: reverse dosimetry using Bayesian inference, Markov chain Monte Carlo simulation, and human biomonitoring data. *Environ. Health Perspect.* **116**, 1040–1046 (2008). doi: 10.1289/ehp.11079
- Mackenzie, P.B.: An improved hybrid Monte Carlo method. Fermilab preprint-89/100-T, March 1989
- Malkus, W.V.R.: Non-periodic convection at high and low Prandtl number. *Mem. Soc. Royal Sci. Leige* **IV**, 125–128 (1972)
- Marcinkiewicz, J.: Sur une propriété de la loi de Gauss (On a property of Gauss's law). *Math. Z.* **44**, 612–618 (1939)
- Maybhate, A., Amritkar, R.E.: Use of synchronization and adaptive control in parameter estimation from a time series. *Phys. Rev. E* **59**, 284 (1999)
- Meliza, C.D., Kostuk, M., Huang, H., Nogaret, A., Abarbanel, H.D.I., Margoliash, D.: Dynamical state and parameter estimation validated by prediction of experimental membrane voltages for conductance-based models of individual neurons. To be submitted to *Neuron* Winter, 2013
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087 (1953)
- Mueller, T.G., Noykova, N., Gyllenberg, M., Timmer, J.: Parameter identification in dynamical models of anaerobic waste water treatment. *Math. Biosci.* **177/178**, 147–60 (2002). doi: 10.1016/S0025-5564(01)00098-0
- Neal, R.M.: Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 25 Sept 1993
- Nijmeijer, H., Mareels, I.M.Y.: An observer looks at Synchronization. *IEEE Trans. Circuit. Syst. I* **44**, 882–890 (1997)
- Northrop, T.G.: *The Adiabatic Motion of Charged Particles*. Interscience, New York (1963)
- Onsager, L., Machlup, S.: Fluctuations and irreversible processes. *Phys. Rev.* **91**, 1505–1512 (1953)

- Oseledec, V.I.: A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.* **19**, 197–231 (1968)
- Panning, T.D., Watson, L.T., Allen, N.A., Chen, K.C., Shaffer, C.A., Tyson, J.J.: Deterministic parallel global parameter estimation for a model of the budding yeast cell cycle. *J. Glob. Optim.* **40**, 719–738 (2008). doi: 10.1007/s10898-007-9273-7
- Parlitz, U.: Estimating model parameters from time series by auto-synchronization. *Phys. Rev. Lett.* **76**, 1232–1235 (1996)
- Parlitz, U., Yu, D.: Inferring network connectivity by delayed feedback control. *PLoS One* **6**, e24333 (2011)
- Parlitz, U., Junge, L., Kocarev, L.: Synchronization based parameter estimation from time series. *Phys. Rev. E* **54**, 6253–6529 (1996)
- Pawula, R.F.: Approximation of the linear Boltzmann equation by the Fokker-Planck equation. *Phys. Rev.* **162**, 186–188 (1967)
- Pawula, R.F.: Generalizations and extensions of the Fokker-Planck-Kolmogorov equations. *IEEE Trans. Inf. Theor.* **13**, 33–41 (1967)
- Pedlosky, J.: *Geophysical Fluid Dynamics*, 2nd edn. Springer, New York (1987) ISBN: 0-387-90368-2
- Pesin, Y.B.: Characteristic Lyapunov exponents and smooth ergodic theory. *Russ. Math. Surv.* **32**, 55–114 (1977)
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. Cambridge University Press, Cambridge (2007). ISBN: 9780521880688
- Prinz, A.A., Bucher, D., Marder, E.: Similar network activity from disparate circuit parameters. *Nature Neurosci.* **7**, 1345–1352 (2004)
- Purcell, E.M.: *Electricity and Magnetism (Berkeley Physics Course)*, vol. 2. McGraw Hill, New York (1965)
- Pythian, R.: The functional formalism of classical statistical dynamics. *J. Phys. A: Math. Gen.* **10**, 777–789 (1977)
- Quinn, J.C., Abarbanel, H.D.I.: State and parameter estimation using Monte Carlo evaluation of path integrals. *Quart. J. R. Meteorol. Soc.* **136**, 1855–1867 doi: 10.1002/qj.690 (2010)
- Quinn, J.C., Abarbanel, H.D.I.: Data assimilation using a GPU accelerated path integral Monte Carlo approach. *J. Comp. Phys.* **230**, 8168–8178, (2011)
- Quinn, J.C., Bryant, P.H., Creveling, D.R., Klein, S.R., Abarbanel, H.D.I.: Parameter and state estimation of experimental chaotic systems using synchronization. *Phys. Rev. E* **80**, 016201 (2009)
- Rabinovich, M.I., Varona, P., Selverston, A.I., Abarbanel, H.D.I.: Dynamical principles in neuroscience. *Rev. Mod. Phys.* **78**, 1222–1265 (2006)
- Restrepo, J.M.: A path integral method for data assimilation. *Phys. D: Nonlinear Phenom.* **237**, 14–27 (2008). doi: 10.1016/j.physd.2007.07.020
- Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R.: Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* **7**, 483–501 (2006). doi: 10.1186/1471-2105-7-483
- Rubinstein, R.Y., Kroese, D.P.: *Simulation and the Monte Carlo method*, 2nd edn. Wiley, New York (2008)
- Sadournay, R.: The dynamics of finite difference models of the shallow water equations. *J. Atmos. Res.* **32**, 680–689 (1975)
- Sakaguchi, H.: Parameter evaluation from time sequences using chaos synchronization. *Phys. Rev. E* **65**, 02720 (2002)
- Sauer, T., Yorke, J.A., Casdagli, M.: *Embedology*. *J. Stat. Phys.* **65**, 579–616 (1991)
- Schlichting, H., Gersten, K.: *Boundary Layer Theory*, 8th edn. Springer, Berlin (2000)
- Schulz, D.J., Goaillard, J.-M., Marder, E.: Variable channel expression in identified single and electrically coupled neurons in different animals. *Nature Neurosci.* **9**, 356–62 (2006)
- So, P., Ott, E., Dayawansa, W.P.: Observing chaos: Deducing and tracking the state of a chaotic system from limited observation. *Phys. Rev. E* **49**, 2650–2660 (1994)

- Stein, P.S.G., Grillner, S., Selverston, A.I., Stuart, D.G. (eds.): *Neurons, Networks, and Motor Behavior*. MIT Press, Cambridge (1997). ISBN-10: 0-262-19390-6, ISBN-13: 978-0-262-19390-0
- Strang, G.: *Introduction to Applied Mathematics*, p. 175. Wellesley-Cambridge Press, Wellesley (1986)
- Strogatz, S.H.: *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*, pp. 302–311. Westview Press, Perseus Books Publishing LLC., Cambridge (1994)
- Sugar, R., Blankenbecler, R.: Variational upper and lower bounds for multichannel scattering. *Phys. Rev.* **136**, B472–B491 (1964)
- Swameye, I., Müller, T.G., Timmer, J., Sandra, U., Klingmüller U.: Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. of the Nat. Acad. of Sci. USA* **100**, 1028–33 (2003). doi: 10.1073/pnas.0237333100
- Swensen, A.M., Bean, B.P.: Robustness of burst firing in dissociated purkinje neurons with acute or long-term reductions in sodium conductance. *J. Neurosci.* **25**, 3509–3520 (2005)
- See the extensive compilation of models, currents, and their mathematical representation in senselab.med.yale.edu/modeldb/showmodel.asp?model=83319
- Toth, B.A.: Dynamical estimation of neuron and network properties. *SIAG/OPT Views-and-News* **21**, 1–8 (2010)
- Toth, B.A., Kostuk, M., Meliza, C.D., Margoliash, D., Abarbanel, H.D.I.: Dynamical estimation of neuron and network properties I: variational methods. *Biol. Cybern.* **105**, 217–237 (2011). doi: 10.1007/s00422-011-0459-1
- Vallis, G.K.: *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*. Cambridge University Press, New York (2006)
- Voss, H., Timmer, J., Kurths, J.: Nonlinear system identification from uncertain and indirect measurements. *Int. J. Bifurcat. Chaos* **14**, 1905–1933 (2004)
- Wächter, A., Biegler, L.T.: On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Program.* **106**(1), 25–57 (2006)
- West, B.J., Bulsara, A.R., Lindenberg, K., Seshadri, V., Schuler, K.E.: Stochastic processes with non-additive fluctuations: I. Itô and Stratonovich calculus and the effects of correlations. *Physica* **97A** **21**, 211–233 (1979)
- Wunsch, C., Heimbach, P.: Practical global ocean state estimation. *Physica D* **230**, 197–208 (2007). doi:10.1016/j.physd.2006.09.040
- Xiong, R., Wissmann, P.J., Gallivan, M.A.: An extended Kalman filter for in situ sensing of yttria-stabilized zirconia in chemical vapor deposition. *Comput. Chem. Eng.* **30**, 1657–1669 (2006). doi: 10.1016/j.compchemeng.2006.05.032
- Yang, Z., Hamrick, J.M.: Optimal control of salinity boundary condition in a tidal model using a variational inverse method. *Estuarine, Coastal Shelf Sci.* **62**, 13–24 (2005). doi: 10.1016/j.ecss.2004.08.003
- Zeigler, H.P., Marler, P. (eds.): *Behavioral Neurobiology of Birdsong*, vol. 1016. *Annals of the New York Academy of Sciences*, New York (2004) [Zeigler (Hunter College of CUNY, New York) and Marler (University of California, Davis, California)]
- Zinn-Justin, J.: *Quantum Field Theory and Critical Phenomena*, 4th edn. Oxford University Press, Oxford (2002)

Index

Symbols

L critical; L_c , 156, 164, 174, 179, 222
4DVar, 90–92, 124, 222, 230

A

Abarbanel, H. D. I., 15–17, 25, 26, 83, 123,
146, 149, 151–153, 163, 166, 181, 185,
223
accepted path, 104–107, 116, 120, 165,
167–169, 171, 188
action, 34, 56, 62, 63, 65, 66, 72, 74–76, 79, 82,
84, 86, 88, 89, 91–94, 97, 98, 100–102,
104–106, 108–111, 113–116, 119, 139,
161, 164–166, 171–173, 179, 180, 187,
222, 223
continuous time, 82
effective, 66, 76, 92–94, 97, 98, 100, 101,
124
loop expansion, 98
residual, 75, 86
action potential, 37, 134, 208, 209, 211
Alexander, F., 83
applied current, 38, 39, 128, 131, 135, 136,
140, 202, 207, 209, 211
attractor, 12, 20, 21, 26, 28, 59, 66, 71, 120,
158, 181, 183, 184, 189
strange, 20, 21, 66, 158

B

Bernstein, I., 83
beta plane, 175
bifurcation, 10, 18–20, 146, 228
birdsong, 200, 201, 218, 229, 232
Blankenbecler, R., 102

C

Carroll, T., 13, 25, 27, 130, 154, 227
Casdagli, M., 224
Caves, C. M., viii
central pattern generator
pyloric, 200
Chandler, D., 80
chaos, 1, 8, 10, 12–20, 22, 24–27, 30, 31,
33–35, 48, 66, 71, 79, 90, 108, 115,
117, 118, 128–130, 136, 137, 145, 146,
149–151, 158, 159, 171, 174, 179–181,
183–186, 188–191, 193–195, 197,
208–211, 213, 215, 217
Chapman, S., 61, 64, 73, 86, 228
Chapman-Kolmogorov equation, 61, 64, 73, 86
Colpitts, E. H., 7, 17–20, 22–25, 28–33, 35, 90,
199, 228
communications, vii, ix, 3, 4, 14, 19, 59, 61,
188
complex, ix, 1–3, 8, 10, 11, 16, 33, 34, 46, 49,
56, 66, 89, 127, 134, 135, 140, 163, 200,
205, 208, 209, 215, 219, 222, 229
complexity, 1, 38
conditional Lyapunov exponent, 14, 25–27, 30,
115, 151–157, 165, 166, 179, 185
conditional mutual information, 58, 59, 61,
63–66, 70, 72, 76, 78, 82, 84, 113, 114,
119, 166, 172, 173
conditional probability, viii, 54, 59–61, 63, 64,
70–73, 84, 221
Cornuelle, B., xi
cost function, 11–13, 22–25, 27–35, 41, 42,
44, 47, 88, 129, 130, 147, 148, 150,
157–159, 225
Cox, H., 83
Creveling, D., 227, 231

- cumulant, 92–94
 cumulant generating function, 93–97
- D**
- data assimilation, vii–x, 2, 3, 5, 11, 17, 18, 29, 33, 34, 36, 42, 44, 47, 50, 51, 58, 67–69, 74, 78, 79, 83–86, 90, 94, 106–108, 115, 118–120, 122, 123, 126–129, 131–135, 141, 145, 155, 159, 166, 167, 170–175, 187, 199, 204, 207, 208, 216, 219, 222, 227, 230
- communications, 3
- general formulation, 11
- general problem, 67
- geophysics, 2, 17, 174
- path integral, 34, 107, 108, 118, 120, 122, 123, 145, 159, 170, 173, 219
- statistical, viii–x, 2, 5, 51, 52, 69, 74, 78, 83–85, 90, 94, 123, 126, 127, 145, 155, 159, 175, 187, 193, 199, 214, 219, 227
- without data, 52, 84
- Debye, P., ix
- depolarized, 136, 137
- deterministic dynamics, 33, 52, 57, 59, 67, 68, 73, 75, 84, 91, 95, 98, 102, 106, 112, 119, 123, 162–167, 171, 223
- direct method, 36, 88, 89, 158, 207
- distribution function, viii, 69, 75, 123, 163
- dynamical systems, vii, 3, 4, 12, 15, 16, 25, 26, 38, 53, 59, 63, 66, 67, 71, 78, 90, 91, 108, 119, 130, 152, 153, 161, 207, 223, 231
- dynamical variables, 3, 4, 10, 15, 18, 20, 32, 58, 90, 145, 180, 181
- Dyson, F. J., 94–98
- Dyson–Schwinger equations, 94
- nonlinear, 97
- Dyson–Schwinger equations, 95, 96
- E**
- Ebers–Moll model, 19, 29, 31
- Ekman layer, 175, 176, 178, 181–186, 188, 190, 192, 194, 196
- electronic circuit, 7, 16, 46, 145
- entropy, 66, 70–72
- error
- model, 2, 5, 12, 33, 35, 49, 52, 63, 69, 75, 83–87, 89, 91, 95, 101, 106, 107, 109, 111–117, 119–123, 126, 128–130, 132, 135, 139, 141, 145, 146, 157, 162, 164, 166, 171, 187, 207
- observation, 2, 52, 63, 67, 83, 84, 86, 90, 95, 106, 107, 109, 110, 114, 115, 171–173, 187
- estimation, x, 2, 3, 12–20, 22–25, 27–29, 31–35, 41–49, 69, 79, 92, 106–108, 111, 113, 120, 124, 126, 129–134, 136–139, 141, 142, 144, 147, 152, 157–161, 166, 169–171, 173–175, 178, 180, 187, 189–192, 194, 196, 203, 208–212, 214, 216, 218, 222, 224, 225, 227, 231
- bias, 111
- parameter, 2, 3, 13, 15–19, 22, 23, 25, 28, 29, 31, 34, 38, 40, 41, 46–48, 107, 120, 126, 129, 132, 140, 144, 227–231
- state variable, 17–19, 22, 29, 31, 43, 48, 120, 139, 141, 174
- state variables, 2
- Evenson, G., ix
- expectation value, 52, 56, 62, 66, 75, 78, 85, 86, 88, 91–94, 101, 106, 108, 109, 111, 114, 116–118, 121, 124, 161, 168, 188, 190, 192, 194, 196, 221
- conditional, 65, 74, 75, 86, 88, 91, 101, 116, 123, 159, 221
- Eyink, G., 83
- F**
- Fano, R., 58, 59, 61, 62, 70, 84
- Feynman, R., 56, 58
- fluid dynamics, ix, 5, 7, 8, 46, 100, 145, 175–178
- shallow water flow, 190
- Fokker–Planck, 101, 102, 110, 163
- Friedland, B., 83
- functional integral, 81
- G**
- gating variables, 37, 39, 41, 43, 45, 46, 49, 107, 128, 131, 132, 135, 136, 138, 141, 143, 204, 205, 209–212, 214, 216, 218
- Gaussian distribution, 49, 75, 76, 79, 83, 84, 86, 90, 91, 95, 101, 102, 106, 108–110, 113–117, 119–121, 159, 160, 163, 168–173, 187, 231
- kurtosis, 116, 117, 119, 120, 122, 159, 160, 168–173
- skewness, 116, 117, 119, 120, 122, 159, 160, 168–173
- Gentner, T., 199
- geostrophic flow, 178
- Gill, A. E., 175, 178, 180

Gill, P. E., x, 35, 36, 88, 163, 166, 227, 228
 GPU, 104, 105, 107, 111–113, 187, 188
 Gummel-Poon model, 29, 31

H

Hahnloser, R., xi
 Hastings, W. K., 103–105, 107, 110, 111, 116,
 162, 165, 168, 169, 172, 187
 Herz, A., xi
 Hochberg, D., 81
 Hodgkin, A., 7
 Hodgkin-Huxley model, 7, 36–42, 125–128,
 130, 131, 134, 135, 140, 141, 144, 199,
 202–205, 207, 217
 ion current, 127
 Howard, L., 7
 Huxley, A., 7
 HVC, 199–201, 208, 210
 hyperpolarized, 134, 136, 137, 139, 140

I

Illing, L., 8, 9, 199, 229
 instability, ix, 1, 3, 14–16, 25, 27, 28, 50, 71,
 146, 163, 209
 integral representation, viii, ix, 58, 79, 84
 ion current, 37, 127, 201, 204, 210, 212, 214,
 216–219
 IPOPT, 35, 36, 88, 89, 158, 207, 232

J

Jacobian matrix, 25–27, 150, 153, 154
 Jazwinski, A. H., 83
 Jouvett, B., 83

K

Kantz, H., 15–17, 20, 25, 26, 83, 146, 149, 180,
 181, 185, 223, 224, 229
 Kennedy, M. P., 17
 kinetic equations, 37, 107, 127, 128
 Koch, C., 199
 Kolmogorov, A. N., 61, 64, 73, 86, 229, 231
 Kostuk, M., 39, 115, 151–157, 169, 200, 208,
 218, 219, 229

L

Langevin equations, 101–103, 163–165, 179
 Laplace, P. S., ix, 84, 87, 88, 92, 100, 124, 203
 Law, K. J. H., x
 limit cycle, 19, 24, 146

Lindenberg, K., 53, 56, 58, 163, 230, 232
 Lindley, D. V., 62
 linear, ix, 3, 17, 18, 26, 37, 53, 56, 76, 79, 86,
 89, 97, 111, 114, 127, 145, 181, 228,
 230, 231
 local minima, 12, 13, 24, 25, 27, 29, 30, 48,
 148, 150, 151, 162–164, 166, 185
 Lorentzian distribution, 172, 173
 Lorenz, E. N., 8, 115–118, 121, 123, 145,
 147–161, 164–167, 170–173, 230
 Lorenz96 model, 115–118, 121, 123, 145,
 147–161, 164–168, 170–173
 Lyapunov exponent, 14, 15, 25–27, 31, 71, 90,
 149, 151–156, 159–161, 169, 173, 180,
 181, 183, 188–190, 194, 223, 231

M

Machlup, S., 83
 Malkus, W., 7–10, 15, 41, 90, 199, 208, 211
 Marcinkiewicz, J., 79, 98
 Marcinkiewicz-Pawula theorem, 79, 98
 marginal distribution, 58, 74, 78, 116, 124,
 167–169, 221
 Margoliash, D., x, 39, 199–201, 208, 211, 218,
 219
 Markov, 53–55, 57, 58, 61–64, 73, 74, 89, 106,
 223, 228, 229
 Markov chain, 63, 64, 74, 106, 228, 229
 master equation, 56, 79, 84
 maximal conductance, 37–39, 107, 128, 140,
 144, 204, 217
 measurement function, 34, 47, 67–69, 75, 76,
 79, 85, 115, 124, 224, 225
 measurement weighting function, 82
 measurements
 how many?, 126, 222
 how often?, 222
 information, 224
 not enough?, 223
 which ones?, 222
 Meliza, C. D., 39, 200, 208, 211, 218, 219
 membrane voltage, 36, 38–41, 43, 45, 49,
 108–110, 127–129, 131–133, 136–138,
 142, 202, 203, 205, 209–218
 Metropolis, N., 103–105, 107, 110, 111, 116,
 162, 165, 168, 169, 172, 187
 Mindlin, G. B., 201
 model, 2
 completed, x, 2, 7, 11, 16, 44, 69, 78, 211,
 218, 222
 observed system, 122
 parameters, 2, 3, 7, 17–19, 52, 68, 69, 78,
 131, 135, 141, 170, 192, 207, 216, 223

- model (*cont.*)
 validation, 2, 5, 11
 validity, 39
- model error, viii, 22, 35, 63, 75, 82–84, 86, 87, 89, 91, 95, 101, 106, 107, 109, 111–117, 119–123, 126, 128–130, 132, 135, 139, 141, 145, 146, 157, 162, 164, 166, 171, 187, 203, 207, 220, 222, 223
 continuous time, 82
- Monte Carlo, ix, 80, 84, 87, 103–108, 110–112, 116, 119, 124, 145, 146, 159, 161–163, 165–174, 187, 222
 hybrid, 103
- mutual information, 58–63, 65, 77, 82, 84, 172
 average, 59, 61, 77
- N**
- NaKL neuron model, 39–49, 106, 108–112, 127–136, 139–144
- NaKLh neuron model, 45, 48, 49, 134–144
- Navier-Stokes equations, 176
- Nernst, E., 37, 127
- network, x, 1, 7, 36–38, 126, 132, 138, 200–202, 210, 213, 214, 218, 232
- neurobiology, ix, 5, 7, 36, 46, 90, 126, 132, 145, 199, 200, 203
- Newton, I., 3, 90
- Nike method, 124
- Nogaret, A., 200, 208, 211, 218, 219
- noise, viii, 33, 40, 57, 58, 63, 66, 69, 76, 91, 102, 106, 108, 110, 113, 115, 116, 118–120, 123, 126, 132–134, 139, 161, 163, 165, 168, 172, 173, 187, 199, 202, 222, 226
 additive, 40, 76, 80, 118, 119, 132–134, 161, 163, 168, 173
- nonlinear, ix, 3, 7, 10, 12, 13, 15–17, 19, 25, 29, 31, 33–36, 41, 46, 50, 56, 63, 68, 71, 79, 89, 90, 97, 101, 129, 133, 146, 171, 200, 223, 227–229, 231, 232
- nonlinear circuit, ix
- numerical estimation, 47, 111, 113
- numerical optimization, 16, 17, 22, 24, 28, 29, 31–36, 41, 42, 44, 49, 88, 94, 124, 129, 207
 direct method, 36, 88
- numerical simulation, 5, 17, 29
- 84, 85, 88, 92, 106, 109–111, 116–121, 130, 131, 138, 146, 158, 159, 169, 175, 187–189, 191, 192, 194, 196, 207, 208, 210, 212, 214, 216, 218, 221, 223, 225
- observations, 2
 sufficient, 2
- observed state variables, vii, 1, 2, 5, 11, 14–17, 30–32, 35, 38, 44, 45, 50, 58, 67–69, 73, 78, 104, 109, 110, 115, 117–122, 126, 129, 130, 133, 141, 156, 157, 159, 164–166, 168, 169, 171–174, 182, 184, 185, 187–192, 194, 196, 203, 209–219, 223, 224
- observer theory, 3, 230
- Onsager, L., 83
- Oseledec, V. I., 26, 27, 149, 151, 153, 154, 179
- Ott, E., 13
- P**
- parallel processing, ix, 89, 104, 105, 112, 113, 124, 222, 231
- Parlitz, U., x, 13, 223, 224, 231
- path integral, viii–x, 5, 34, 47, 52, 55, 56, 58, 66, 77–80, 83, 85, 88, 89, 92, 94, 98, 99, 106, 108, 109, 111, 112, 115, 117–124, 126, 139, 145, 146, 159, 160, 162–167, 170–174, 187–189, 203, 214, 221, 222, 226
 continuous time, 80
 data assimilation, viii, x, 34, 107, 108, 115, 118–120, 123
 Monte Carlo, 106, 108, 112, 166, 167, 170
- Pawula, R. F., 79, 98
- Pecora, L., 13, 25, 27, 130, 154, 227
- Pedlosky, J., 175, 178, 180, 231, 232
- Pesin, Y. B., 71, 231
- Poincaré section, 20
- polynomial chaos, 79
- predict, viii, 2–4, 7, 10, 11, 15, 16, 18, 19, 29, 31–36, 38, 42, 44–47, 50, 68, 69, 78, 79, 85, 100, 106, 107, 109–111, 116–118, 120, 128, 137–139, 141, 145, 146, 148, 157, 159–161, 166–171, 173–175, 178, 180, 181, 186–197, 211, 212, 214, 216, 218, 222, 223, 225
- prediction, 22, 138, 209, 214
- probability density, 53, 54
- probability distribution, viii, 52–58, 61, 63, 66, 69, 70, 72, 79, 107, 163, 221
 conditional, 59, 61, 70, 72, 221
- Pythian, R., 83
- objective function, 34–36, 47, 88
- observation window, viii, 5, 15, 16, 29–35, 38, 41, 42, 44–46, 48, 67–69, 72, 74, 78,

Q

Quinn, J. C., 17, 29, 104, 113, 115, 116, 199, 231

R

Rabinovich, M. I., x, 228, 230, 231
 regularization, ix, 16, 25, 27–33, 35, 44, 48, 50, 129, 130, 157, 205, 207, 227
 reversal potential, 39, 107, 127, 128, 204, 206
 RMS, 74, 106–111, 116–118, 120, 121, 133, 159, 161, 168, 170, 171, 173, 174, 180, 182–188, 190, 192, 194, 196
 Rosenbluth, M. N., 103–105, 107, 110, 111, 116, 162, 165, 168, 169, 172, 187
 Runge-Kutta, 11, 40, 106, 108, 154, 166, 168, 170, 171, 173
 fourth order, 40, 106, 170, 171, 173

S

saddle path, ix, xi, 87–89, 92, 94, 103, 107, 110, 124, 132, 203, 207, 222
 saddle point, ix, 87, 124, 203
 Sauer, T., 224
 Schreiber, T., 15–17, 20, 25, 26, 83, 146, 149, 180, 181, 185, 223, 224, 229
 Schwinger, J., 94–98
 self-consistent errors, 112–114, 117
 Selverston, A. I., x, 200, 231, 232
 serendipity, 178, 199
 shallow water equations, 125, 145, 174–176, 178–188, 190, 192, 194, 196, 231
 Shannon, C., 70, 84
 SNOPT, 35, 36, 88, 89, 228
 song system, 200, 201, 218
 spike, 37, 134, 208–211
 state space, ix, 1–5, 7, 15, 17–20, 29, 30, 35, 39, 41, 49, 52, 53, 62, 63, 65, 66, 68, 71, 73, 75, 78, 83–85, 88, 91, 92, 103, 106, 110–113, 119, 120, 127, 128, 131, 132, 135, 136, 138, 154, 158, 162, 170, 202, 207, 209, 219, 221–223
 state variable, 82
 state variables, 11, 13, 80
 stationary path, 86, 87, 90, 92, 94, 203
 statistical physics, viii, 5, 80, 83, 92, 93, 95, 98
 stimulus, 38, 126, 129, 131, 132, 135–141, 208
 Strogatz, S., 8, 10, 20, 232
 Stuart, A. M., 83
 Sugar, R. L., 102
 synaptic plasticity, 201

synchronization, 4, 12–15, 19, 23–25, 27–30, 50, 115, 129, 130, 146–157, 163–166, 174, 175, 179, 182–187, 189–192, 194, 196, 207, 222, 223
 synchronization error, 14, 23, 147–149, 164, 165, 179, 182–187
 synchronization manifold, 25, 27–30, 50, 115, 129, 130, 149, 151–157, 163, 166, 179, 185, 223
 syrinx, 201

T

time delay, 53, 223, 224
 time delay coordinates, 224
 time delay embedding, 223
 time delay embedding space, 224
 time derivatives
 discrete time approximation, 224
 Toth, B., 39, 89, 227
 transistor, 17–20, 29, 31, 35
 bipolar, 17–20, 29, 31, 35
 transition probability, 54–57, 61, 63, 64, 73, 75, 78, 86, 89, 91, 112, 114, 119, 123, 162, 187, 222, 223
 twin experiments, ix, x, 5, 7, 11, 12, 16, 17, 34, 40–44, 46, 49, 106, 108, 111, 115, 120, 125–128, 131, 132, 135–139, 141, 146, 173, 174, 179, 190, 192, 193, 199, 201, 203, 219, 222

U

unobserved state variables, viii, x, 2, 3, 10, 11, 15–18, 22, 29, 31, 33, 35, 38–41, 43, 45–47, 49, 50, 67–69, 78, 88, 109, 117, 118, 120, 122, 126, 128, 130–133, 139, 141, 146, 148, 152, 157–161, 168–171, 173, 174, 185, 188–192, 194, 196, 203, 210, 212, 216, 223–225

V

validation, 2, 11, 22, 31, 32, 39, 126, 137–139, 141, 174, 212
 Vallis, G., 175, 178, 232
 variational principle, 28, 31, 33, 34, 36, 47, 49, 84, 88–90, 92, 98, 102, 103, 107, 128, 129, 132, 135, 139, 141, 145, 146, 157, 162, 165, 207, 227, 232

Varona, P., [x](#), [231](#)

voltage-gated channel, [38](#), [39](#), [134](#), [203](#), [205](#)

W

waterwheel, [7](#), [8](#), [10](#), [11](#), [13](#), [15](#), [16](#), [41](#), [90](#), [199](#),
[208](#), [210](#)

West, B., [53](#), [56](#), [58](#), [230](#), [232](#)

Y

Yorke, J., [224](#)

Z

Zinn-Justin, J., [55](#), [56](#), [58](#), [81](#), [92](#), [95](#), [98](#), [100](#),
[124](#), [163](#), [232](#)