Edited by Tim Palmer
and Renate Hagedorn

Predictability of Weather and Climate

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521848824

This page intentionally left blank

Predictability of Weather and Climate

The topic of predictability of weather and climate has advanced significantly over recent years, both through an increased understanding of the phenomena that affect predictability, and through development of techniques used to forecast state-dependent predictability.

This book brings together some of the world's leading experts on predictability of weather and climate. It addresses predictability from the theoretical to the practical, on timescales from days to decades. Topics such as the predictability of weather phenomena, coupled ocean–atmosphere systems and anthropogenic climate change are among those included. Ensemble systems for forecasting predictability are discussed extensively. Ed Lorenz, father of chaos theory, makes a contribution to theoretical analysis with a previously unpublished paper.

This well-balanced volume will be a valuable resource for many years. High-quality chapter authors and extensive subject coverage will make it appeal to people with an interest in weather and climate forecasting and environmental science, from graduate students to researchers.

TIM PALMER is Head of the Probability Forecasting and Diagnostics Division at the European Centre for Medium-Range Weather Forecasts (ECMWF). He has won awards including the American Meteorological Society Charney Award and the Royal Meteorological Society Buchan Award. He is a fellow of the Royal Society, and is co-chair of the Scientific Steering Group of the World Climate Research Programme's Climate Variability and Predictability (CLIVAR) project. He was a lead author of the Intergovernmental Panel on Climate Change Third Assessment Report, and coordinator of two European Union climate prediction projects: PROVOST and DEMETER.

RENATE HAGEDORN is the education officer for the ECMWF research department. She gained her Ph.D. at the Institute for Marine Sciences in Kiel, Germany, where she developed a coupled atmosphere ocean model for the Baltic Sea catchment area. Upon joining the ECMWF she was part of the DEMETER team. More recently she has been working on improving the ECMWF ensemble prediction systems. In conjunction with Tim Palmer and other colleagues, she was awarded the Norbert Gerbier-Mumm International Award 2006 by the World Meteorological Organization.

Predictability of Weather and Climate

Edited by

Tim Palmer and Renate Hagedorn

European Centre for Medium-Range Weather Forecasts, Reading, UK

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521848824

First published in print format 2006

ISBN-13 978-0-511-22300-6 eBook (Adobe Reader)
ISBN-10 0-511-22300-5 eBook (Adobe Reader)

ISBN-13 978-0-521-84882-4 hardback
ISBN-10 0-521-84882-2 hardback

Contents

The colour plates are situated between pages 364 and 365.

Contributors

Myles Allen Department of Atmospheric Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

David Anderson ECMWF, Shinfield Park, Reading RG2 9AX, UK

Roberto Buizza ECMWF, Shinfield Park, Reading RG2 9AX, UK

Arun Chakraborty Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

H.-R. Chang School of Earth and Atmospheric Sciences, Atlanta, GA, USA

Susanna Corti Institute of Atmospheric Sciences and Climate, ISAC-SNR, Bologna, Italy

Francisco J. Doblas-Reyes ECMWF, Shinfield Park, Reading RG2 9AX, UK

Martin Ehrendorfer Inst. für Meteorologie und Geophysik, Universität Innsbruck, Innrain 52, A-6020 Innsbruck, Austria

Brian Farrell Harvard University, Division of Engineering and Applied Sciences, Pierce Hall 107d, Oxford Street Mail Area H0162, Cambridge, MA 02138, USA

David Frame Department of Physics, University of Oxford, UK

R. Grossman Colorado Research Associates, Boulder, CO, USA

Renate Hagedorn ECMWF, Shinfield Park, Reading RG2 9AX, UK

Thomas Hamill Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO, USA

T. Hopson Program in Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA

Brian Hoskins Department of Meteorology, University of Reading, 2 Early Gate, Whiteknights, Reading RG6 6AH, UK

C. Hoyos School of Earth and Atmospheric Sciences, Atlanta, GA, USA

Brian Hunt Chaos Group, University of Maryland, College Park, MD, USA

Petros Ioannou Section of Astrophysics, Astronomy and Mechanics, Department of Physics, Panepistimiopolis, Zografos 15784, Athens, Greece

Fei-Fei Jin Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

Eugenia Kalnay Department of Meteorology, University of Maryland, 3431 Computer and Space Sciences Building, College Park, MD 20742-2425, USA

Jamie Kettleborough Space Science and Technology Department, Rutherford Appleton Laboratory, Didcot, Oxon, UK

J. L. Kinter III Center for Ocean-Land-Atmosphere Studies, Calverton, MD, USA

Tiruvalam Krishnamurti Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

T. S. V. Vijaya Kumar Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

Fred Kucharski Abdus Salam International Centre for Theoretical Physics, Trieste

François Lalaurette Ecole Nationale de la Météorologie, Av. G. Coriolis, 31057 Toulouse, France

Mojib Latif IfM-Geomar, Düsternbrooker Weg 20, 24105 Kiel, Germany

Edward Lorenz Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Chiara Marsigli ARPA-SIM, Bologna, Italy

Franco Molteni Physics of Weather and Climate Dynamics, The Abdus Salam International Centre for Theoretical Physics, PO Box 586, I-34100 Trieste, Italy

Andrea Montani ARPA-SIM, Bologna, Italy

Ken Mylne Met Office, FitzRoy Road, Exeter EX1 3PB, UK

Fabrizio Nerozzi ARPA-SIM, Bologna, Italy

Edward Ott Chaos Group, University of Maryland, College Park, MD, USA

Tiziana Paccagnella ARPA-SIM, Bologna, Italy

Tim Palmer ECMWF, Shinfield Park, Reading RG2 9AX, UK

Wonsum Park Max-Planck-Institut für Météorologie, Hamburg, Germany

Goðrún Nína Petersen University of Reading, UK

Holger Pohlmann Max-Planck-Institut für Meteorologie, Hamburg, Germany

David Richardson ECMWF, Shinfield Park, Reading RG2 9AX, UK

Jagadish Shukla George Mason University, Institute of Global Environment and Society Inc, 4041 Powder Mill Road, Suite 302 Calverton, MD 20705-3106, USA

Adrian Simmons ECMWF, Shinfield Park, Reading RG2 9AX, UK

Leonard Smith OCIAM Mathematical Institute, 24–29 St Giles', Oxford OXl 3LB, UK

David Stainforth Department of Physics, University of Oxford, UK

Lydia Stefanova Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

A. Subbiah Asian Disaster Preparedness Centre, Bangkok, Thailand

Istvan Szunyogh University of Maryland, College Park, MD, USA

Olivier Talagrand Laboratoire de Météorologie Dynamique, Paris, France

Alan J. Thorpe Department of Meteorology, University of Reading, 2 Early Gate, Whiteknights, Reading RG6 6AH, UK

Stefano Tibaldi Servizio Meteorologico Regionale, ARPA Emilia-Romagna, Viale Silvani, 640122 Bologna, Italy

Axel Timmermann Department of Oceanography, University of Hawaii at Manoa, 1000 Pope Road, Marine Sciences Building, Honolulu, HI 96822, USA

Zoltan Toth Environmental Modeling Center, NCEP, NWS/NOAA, Washington DC 20233, USA

Gerald van der Grijn ECMWF, Shinfield Park, Reading RG2 9AX, UK

Duane Waliser Jet Propulsion Laboratory, MS 183-505, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Peter Webster School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332-0340, USA

Won-Tae Yun Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520, USA

Yuejian Zhu National Centers for Environmental Prediction, Washington DC, USA

Preface

In his biography of the great twentieth-century theoretical physicist Richard Feynman, Gleick (1993) writes: ‘He (Feynman) believed in the primacy of doubt, not as a blemish on our ability to know, but as the essence of knowing’. Feynman’s philosophy applies as much to weather and climate forecasting as to fundamental physics, as made explicit by Tennekes *et al.* (1987) when they wrote: ‘no forecast is complete without a forecast of forecast skill’.

The estimation of uncertainty in weather and climate prediction is encapsulated in the word ‘predictability’. If something is said to be predictable, then presumably it can be predicted! However, initial conditions are never perfect and neither are the models used to make these predictions. Hence, the predictability of the forecast is a measure of how these inevitable imperfections leave their imprint on the forecast. By virtue of the non-linearity of the climate, this imprint varies from day to day, just as the weather itself varies; predictability is as much a climatic variable as rainfall, temperature or wind.

Of course, it is one thing to talk about predictability as if it were just another climatic variable; it is another thing to estimate it quantitatively. The predictability of a system is determined by its instabilities and non-linearities, and by the structure of the imperfections. Estimating these instabilities, non-linearities and structures provides a set of tough problems, and real progress requires sophisticated mathematical analysis on both idealised and realistic models.

However, the big world out there demands forecasts of the weather and the climate: is it going to rain tomorrow, will the Arctic ice cap melt by the end of the century? The man in the street wanting to know whether to bring his umbrella to work, or the politician looking for advice on formulating her country’s strategy on climate change, cannot wait for the analysis on existence or otherwise of heteroclinic state-space orbits to be finalised! The difference between the real world of prediction, and the more aesthetic world of predictability has been perfectly encapsulated by one of the pioneers of the subject, Kiku Miyakoda, who said: 'Predictability is to prediction as romance is to sex!'. Oscar Wilde, who wrote: 'The very essence of romance is uncertainty!', might well have approved.

However, as we enter the twenty-first century, is this still a fair characterisation? We would argue not! In the last decade, the romantic world of predictability has collided head-on with the practical world of prediction. No longer do operational centres make forecasts without also estimating forecast skill – whether for predictions one hour ahead or one century ahead. This change has come in the last few years through the development of ensemble forecast techniques made practical by mind-boggling developments in high-performance computer technology.

In late 2002, the European Centre for Medium-Range Weather Forecasts (ECMWF) held a week-long seminar on the topic of Predictability of Weather and Climate. A subtheme, borrowing from Kiku Miyakoda's aphorism, was to celebrate the 'reconciliation of romance and sex'! World leaders in the field of predictability of weather and climate gave pedagogical presentations covering the whole range of theoretical and practical aspects on weather and climate timescales, i.e. from a few hours to a century. It was decided, as this was a sufficiently landmark meeting and the presentations sufficiently comprehensive, that it was worth publishing the proceedings for the benefit of the larger scientific community. During 2004 and 2005 authors were asked to expand and update their presentations.

In fact there is one exception to this strategy. One of the greatest pioneers of the subject is Ed Lorenz – his prototypical model of chaos spawned a revolution, not only in meteorology, but in mathematics and physics in general. Ed was unable to come to the 2002 meeting, but a few years earlier had given a presentation at ECMWF on what has become known as the Lorenz-1996 model. This paper is widely cited, but has never been published externally. We decided it would be proper to acknowledge Ed's unrivalled contribution to the field of weather and climate predictability by publishing his 1996 paper in this volume.

Lorenz's contribution is one of the introductory chapters on predictability where both general and specific theoretical/mathematical aspects of predictability theory are discussed. These chapters are followed by contributions on data assimilation methods. The next chapters represent a journey through the predictability of different timescales and different phenomena. The link to real-world applications is made by discussing important developments in operational forecast systems, presenting methods to diagnose and improve forecast systems, and finally giving examples utilising predictability in decision-making processes.

We would like to acknowledge the help of Anabel Bowen, Rob Hine, Els Kooij-Connally, and Matt Lloyd during all stages of the production of this book. Last but

not least, we would like to thank ECMWF for initiating and supporting the seminar on which the contributions of this book are based.

References

Gleick, J. (1992) *Genius: The Life and Science of Richard Feynman*. Pantheon.

Tennekes, H., A. P. M. Baede and J. D. Opsteegh (1987) Forecasting forecast skill. In *Proceedings, ECMWF Workshop on Predictability*. ECMWF, Reading, UK.

1

Predictability of weather and climate: from theory to practice

T. N. Palmer
European Centre for Medium-Range Weather Forecasts, Reading

1.1 Introduction

A revolution in weather and climate forecasting is in progress, made possible by theoretical advances in our understanding of the predictability of weather and climate on the one hand, and by the extraordinary developments in supercomputer technology on the other. Specifically, through ensemble prediction, whose historical development has been documented by Lewis (2005), weather and climate forecasting is set to enter a new era, addressing quantitatively the prediction of weather and climate risk in a range of commercial and humanitarian applications. This chapter gives some background to this revolution, with specific examples drawn from a range of timescales.

1.2 Perspectives on predictability: theoretical and practical

Predictions of weather and climate are necessarily uncertain; our observations of weather and climate are uncertain and incomplete, the models into which we assimilate this data and predict the future are uncertain, and external effects such as volcanoes and anthropogenic greenhouse emissions are also uncertain. Fundamentally, therefore, we should think of weather and climate prediction in terms of equations whose basic prognostic variables are probability densities $\rho(X, t)$, where X denotes

Predictability of Weather and Climate, ed. Tim Palmer and Renate Hagedorn. Published by Cambridge University Press.

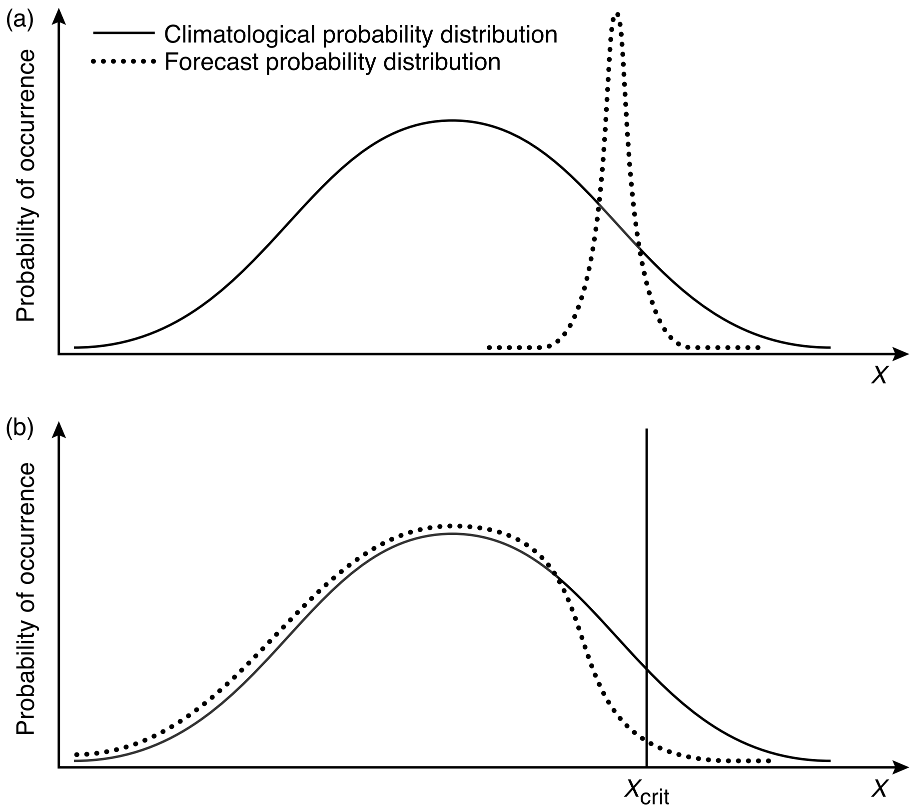

Figure 1.1 Schematic illustration of the climatological probability distribution of some climatic variable X (solid line) and a forecast probability distribution (dotted line) in two different situations. The forecast probability distribution in (a) is obviously predictable. In a theoretical approach to predictability, $\rho(X, t) - \rho_C(X)$ in (b) may not be significantly different from zero overall. However, considered more pragmatically, the forecast probability distribution in (b) can be considered predictable if the prediction that it is unlikely that X will exceed X_{crit} can influence decision-makers.

some climatic variable and t denotes time. In this way, $\rho(X, t)dV$ represents the probability that, at time t, the true value of X lies in some small volume dV of state space. Prognostic equations for ρ, the Liouville and Fokker–Planck equations, are described in Ehrendorfer (this volume). In practice these equations are solved by ensemble techniques, as described in Buizza (this volume).

The question of whether or not X is predictable depends on whether the forecast probability density $\rho(X, t)$ is sufficiently different from some prior estimate $\rho_C(X)$, usually taken as the climatological probability density of X. What do we mean by 'sufficiently different'? One could, for example, apply a statistical significance test to the difference $\rho(X, t) - \rho_C(X)$. On this basis, the hypothetical forecast probability distribution shown as the dotted curve in Figure 1.1(a) is certainly predictable;

but the forecast probability distribution shown in Figure 1.1(b) may well not be predictable.

However, this notion of predictability is a rather idealised one and takes no account of how $\rho(X, t)$ might be used in practice. In a more pragmatic approach to predictability, one would ask whether $\rho(X, t)$ is sufficiently different from $\rho_C(X)$ to influence decision-makers. For example, in Figure 1.1, an aid agency might be interested only in the right-hand tail of the distribution, because disease A only becomes prevalent when $X > X_{\text{crit}}$. On the basis of Figure 1.1(b), the agency may decide to target scarce resources elsewhere in the coming season, since the forecast probability that $X > X_{\text{crit}}$ is rather low.

These two perspectives on the problem of how to define predictability reflect the evolving nature of the study of predictability of weather and climate prediction; from a rather theoretical and idealised pursuit to one which recognises that quantification of predictability is an essential part of operational activities in a wide range of applications. The latter perspective reflects the fact that the full economic value of meteorological predictions will only be realised when quantitatively reliable flow-dependent predictions of weather and climate risk are achievable (Palmer, 2002).

The scientific basis for ensemble prediction is illustrated in Figure 1.2, based on the famous Lorenz (1963) model. Figure 1.2 shows that the evolution of some isopleth of $\rho(X, t)$ depends on starting conditions. This is a consequence of the fact that the underlying equations of motion

$$\dot{X} = F[X] \tag{1.1}$$

are non-linear, so that the Jacobian dF/dX in the linearised equation

$$\frac{d\,\delta X}{dt} = \frac{dF}{dX}\,\delta X \tag{1.2}$$

depends at least linearly on the state X about which Equation (1.1) is linearised. As such, the so-called tangent propagator

$$M(t, t_0) = \exp \int_{t_0}^{t} \frac{dF}{dX}\, dt' \tag{1.3}$$

depends on the non-linear trajectory $X(t)$ about which the linearisation is performed. Hence, the evolved perturbations

$$\delta X(t) = M(t, t_0)\,\delta X(t_0) \tag{1.4}$$

depend not only on $\delta X(t_0)$, but also on the region of phase space through which the underlying non-linear trajectory passes.

It is of interest to note that the formal solution of the Liouville equation, which describes the evolution of $\rho(X, t)$ arising from initial error only (Ehrendorfer, this volume, Eq. (4.49)), can be written using the tangent propagator (for all time in

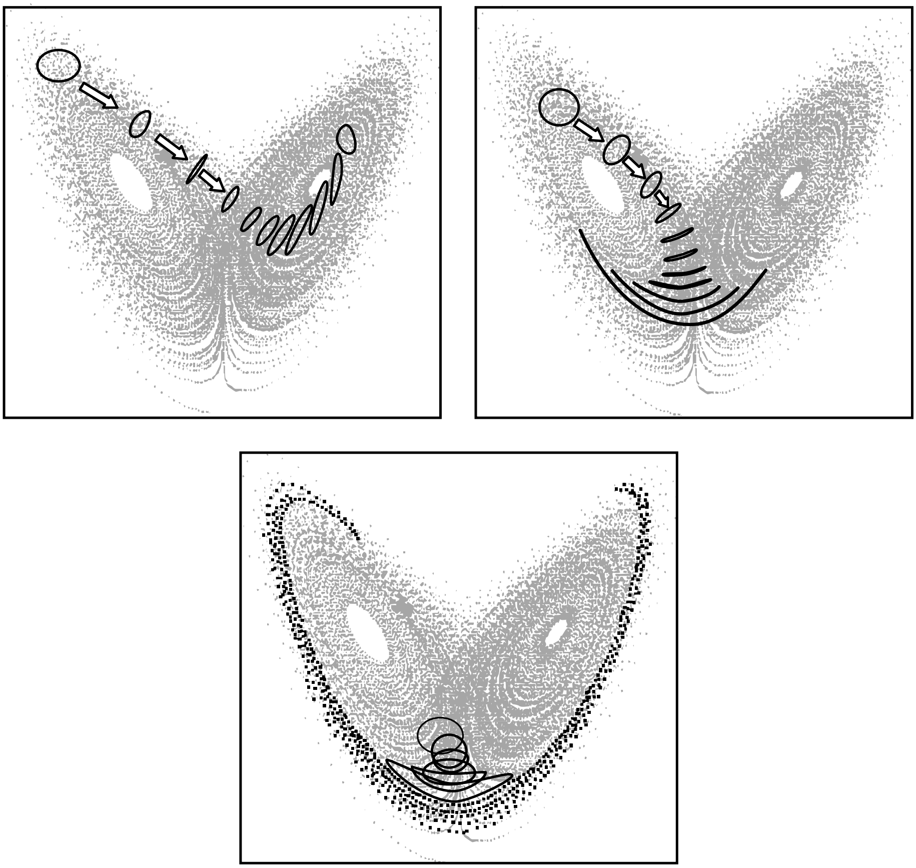

Figure 1.2 Finite time ensembles of the Lorenz (1963) system illustrating the fact that in a non-linear system, the evolution of the forecast probability density $\rho(X, t)$ is dependent on initial state.

the future, not just the time for which the tangent-linear approximation is valid). Specifically

$$\rho(X, t) = \rho(X', t_0)/|\det M(t, t_0)| \tag{1.5}$$

where X' corresponds to the initial state which, under the action of Eq. (1.1), evolves into the state X at time t. Figure 1.2 shows solutions to Eq. (1.5) using an ensemble-based approach.

To illustrate the more practical implications of the fact that $\rho(X, t)$ depends on initial state, I want to reinterpret Figure 1.2 by introducing you to Charlie, a builder by profession, and a golfing colleague of mine! Charlie, like many members of my golf club, takes great pleasure in telling me when (he thinks) the weather forecast has gone wrong. This is mostly done in good humour, but on one particular occasion Charlie was in a black mood. ‘I have only four words to say to you,’ he announced, 'How do I sue?' I looked puzzled. He continued: 'The forecast was for a night-time minimum temperature of five degrees. I laid three thousand square yards of concrete. There was a frost. It's all ruined. I repeat – how do I sue?'

If only Charlie was conversant with Lorenz (1963) I could have used Figure 1.2 to illustrate how in future he will be able to make much more informed decisions about when, and when not, to lay concrete! Suppose the Lorenz equations represent part of an imaginary world inhabited by builders, builders' customers, weather forecasters and lawyers. In this Lorenz world, the weather forecasters are sued if the forecasts are wrong! The weather in the Lorenz world is determined by the Lorenz (1963) equations where all states on the right-hand lobe of the attractor are 'frosty' states, and all states on the left-hand lobe of the attractor are 'frost-free' states. In this imaginary world, Charlie is planning to lay a large amount of concrete in a couple of days' time. Should he order the ready-mix concrete lorries to the site? He contacts the Lorenzian Meteorological Office for advice. On the basis of the ensemble forecasts in the top left of Figure 1.2 he clearly should not – all members of the ensemble predict frosty weather. On the basis of the ensemble forecasts in the bottom left of Figure 1.2 he also should not – in this case it is almost impossible to predict whether it will be frosty or not. Since the cost of buying and laying concrete is significant, it is not worth going ahead when the risk of frost is so large.

How about the situation shown in the top right of Figure 1.2? If we took the patronising but not uncommon view that Charlie, as a member of the general public, would only be confused by a probability forecast, then we might decide to collapse the ensemble into a consensus (i.e. ensemble-mean) prediction. The ensemble-mean forecast indicates that frost will not occur. Perhaps this is equivalent to the real-world situation that got Charlie so upset. Lorenzian forecasters, however, will be cautious about issuing a deterministic forecast based on the ensemble mean, because, in the Lorenz world, Charlie can sue!

Alternatively, the forecasters could tell Charlie not to lay concrete if there is even the slightest risk of frost. But Charlie will not thank them for that either. He cannot wait forever to lay concrete since he has fixed costs, and if he doesn't complete this job, he may miss out on other jobs. Maybe Charlie will never be able to sue, but neither will he bother obtaining the forecasts from the Lorenzian Meterorological Office.

Suppose Charlie's fixed costs are C, and that he loses L by laying concrete when a frost occurs. Then a logical decision strategy would be to lay concrete when the ensemble-based estimate of the probability of frost is less than C/L. The meteorologists don't know Charlie's C/L, so the best they can do is provide him with the full probability forecast, and allow him to decide whether or not to lay concrete.

Clearly the probability forecast will only be of value to Charlie if he saves money using these ensemble forecasts. This notion of 'potential economic value' (Murphy, 1977; Richardson, this volume) is conceptually quite different from the notion of skill (in the meteorological sense of the word), since value cannot be assessed by analysing meteorological variables alone; value depends also on the user's economic parameters.

The fact that potential economic value does not depend solely on meteorology means that we cannot use meteorological skill scores alone if we want to assess whether one forecast system is more valuable than another (e.g. to Charlie). This is relevant to the question of whether it would be better to utilise computer resources to increase ensemble size or increase model resolution. As discussed in Palmer (2002), the answer to this question depends on C/L. For users with small C/L, more value may accrue from an increase in ensemble size (since decisions depend on whether or not relatively small probability thresholds have been reached), whilst for larger C/L more value may accrue from the better representation of weather provided by a higher-resolution model.

In the Lorenz world, Charlie never sues the forecasters for 'wrong' forecasts. When the forecast is uncertain, the forecasters say so, and with precise and reliable estimates of uncertainty. Charlie makes his decisions based on these forecasts and if he makes the wrong decisions, only he, and lady luck, are to blame!

1.3 Why are forecasts uncertain?

Essentially, there are three reasons why forecasts are uncertain: uncertainty in the observations used to define the initial state, uncertainty in the model used to assimilate the observations and to make the forecasts, and uncertainty in 'external' parameters.

Let's consider the last of these uncertainties first. For example, the aerosol content of the atmosphere can be significantly influenced by volcanic eruptions, which are believed to be unpredictable more than a few days ahead. Also, uncertainty in the change in atmospheric CO_2 over the coming decades depends on which nations sign agreements such as the Kyoto protocol.

In principle, perhaps, 'stochastic volcanoes' could be added to an ensemble prediction system – though this seems a rather fanciful idea. Also, uncertainties in humankind's activities can, perhaps, be modelled by coupling our physical climate model to an econometric model. However, we will not deal further with such uncertainties of the 'third kind' but rather concentrate on the first two.

1.3.1 Initial uncertainty

At ECMWF, for example, the analysed state X_{a} of the atmosphere is found by minimising the cost function

$$J(X) = \tfrac{1}{2}(X - X_{\mathrm{b}})^T B^{-1} (X - X_{\mathrm{b}}) + \tfrac{1}{2}(HX - Y)^T O^{-1} (HX - Y) \quad (1.6)$$

where X_{b} is the background state, B and O are covariance matrices for the probability density functions (pdf) of background error and observation error, respectively, H is

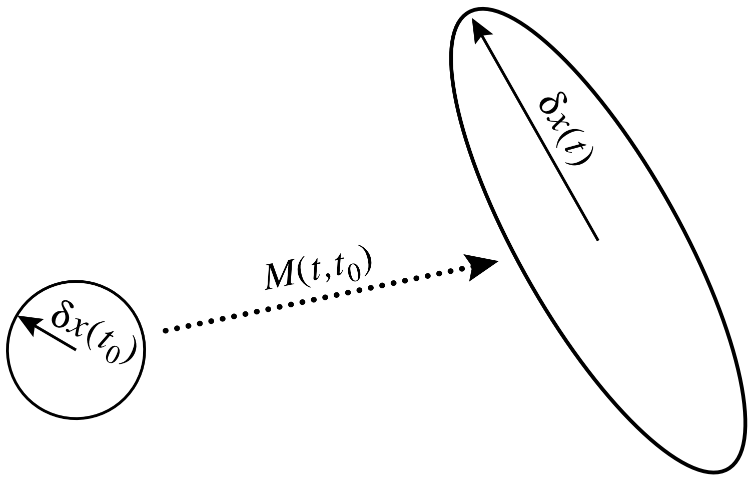

Figure 1.3 Isopleths of probability that the region enclosed by the isopleths contains truth at initial and forecast time. The associated dominant singular vector at initial and final time is also shown.

the so-called observation operator, and Y denotes the vector of available observations (e.g. Courtier *et al.*, 1998). The Hessian

$$\nabla\nabla J = B^{-1} + H^T O^{-1} H \equiv A^{-1} \tag{1.7}$$

of J defines the inverse analysis error covariance matrix.

Figure 1.3 shows, schematically, an isopleth of the analysis error covariance matrix, and its evolution under the action of the tangent propagator M (see Eqs. 1.3 and 1.4). The vector pointing along the major axis at forecast time corresponds to the leading eigenvector of the forecast error covariance matrix. Its pre-image at initial time corresponds to the leading singular vector of M, determined with respect to unit norm in the metric given by A. The singular vectors of M correspond to the eigenvectors of $M^T M$ in the generalised eigenvector equation

$$M^T M\, \delta x(t_0) = -\lambda A^{-1} \delta x(t_0). \tag{1.8}$$

Given pdfs of uncertainty based on Eq. (1.6), we can in principle perform a Monte Carlo sampling of the Hessian-based initial pdf and produce an ensemble forecast system based on this initial sampling.

There are three reasons for not adopting this strategy.

Firstly, there is the so-called 'curse of dimensionality'. The state space of a weather prediction model has about 10^7 dimensions. Many of these dimensions are not dynamically unstable (i.e. are not associated with positive singular values). In this sense, a random sampling of the initial probability density would not be a computationally efficient way of estimating the forecast probability density. This point was made explicitly in Lorenz's analysis of his 28-variable model (Lorenz, 1965):

> If more realistic models . . . also have the property that a few of the eigenvalues of MM^{T} are much larger than the remaining, a study based upon a small ensemble of initial errors should . . . give a reasonable estimate of the growth rate of random error. . . . It would appear, then, that the best use could be made of computational time by choosing only a small number of error fields for superposition upon a particular initial state.

Studies of realistic atmospheric models show that the singular values of the first 20–30 singular vectors are indeed much larger than the remainder (Molteni and Palmer, 1993; Buizza and Palmer, 1995; Reynolds and Palmer, 1998).

The second reason for not adopting a Monte Carlo strategy is that in practice Eq. (1.6) only provides an estimate of part of the actual initial uncertainty; there are other sources of initial uncertainty that are not well quantified – what might be called the 'unknown unknowns'. Consider the basic notion of data assimilation: to assimilate observations that are either made at a point or over a pixel size of kilometres into a model whose smallest resolvable scale is many hundreds of kilometres (bearing in mind the smallest resolvable scale will be many times the model grid). Now sometimes these point or pixel observations may be representative of circulation scales that are well resolved by the model (e.g. if the flow is fairly laminar at the time the observation is made); on other occasions the observations may be more representative of scales which the model cannot resolve (e.g. if the flow is highly turbulent at the time the observation is made, or if the observation is sensitive to small-scale components of the circulation, as would be the case for humidity or precipitation).

In the latter case, the practice of using simple polynomial interpolation in the observation operator H in Eq. (1.6) to take the model variable X to the site of the observation, is likely to be poor. However, this is not an easily quantified uncertainty– since, ultimately, the uncertainty relates to numerical truncation error in the forecast model (see the discussion below). Similarly, consider the problem of quality control. An observation might be rejected as untrustworthy by a quality-control procedure if the observation does not agree with its neighbours and is different from the background (first-guess) field. Alternatively, the observation might be providing the first signs of a small-scale circulation feature, poorly resolved by either the model or the observing network. For these types of reason, a Monte Carlo sampling of a pdf generated by Eq. (1.6) is likely to be an underestimate of the true uncertainty.

The third reason for not adopting a Monte Carlo strategy is not really independent of the first two, but highlights an issue of pragmatic concern. Let us return to Charlie, as discussed above. Charlie is clearly disgruntled by the occasional poor forecast of frost, especially if it costs him money. But just imagine how much more disgruntled he would be, having invested time and money to adapt his decision strategies to a new weather risk service based on the latest, say, Multi-Centre Ensemble Forecast System, if no member of the new ensemble predicts severe weather, and severe

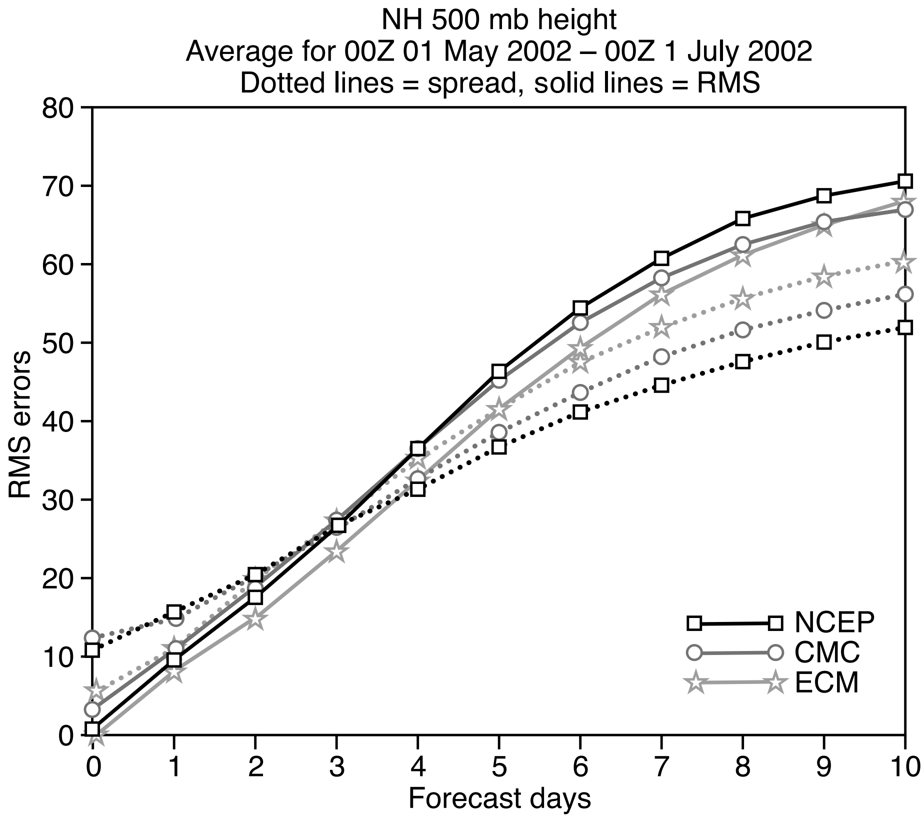

Figure 1.4 May–July 2002 average root-mean-square (rms) error of the ensemble-mean (solid lines) and ensemble standard deviation (dotted lines) of the ECMWF, NCEP and MSC ensemble forecast systems. Values refer to the 500 hPa geopotential height over the Northern Hemisphere latitudinal band 20–80 N. From Buizza *et al.* (2003, 2005).

weather occurs! Just one failure of this sort will compromise the credibility of the new system.

To take this into account, a more conservative approach to sampling initial perturbations is needed, conservative in the sense of tending towards sampling perturbations that are likely to have significant impact on the forecast.

For these three reasons (together with the fact that instabilities in the atmosphere are virtually never of the normal-mode type: Palmer, 1988; Molteni and Palmer, 1993; Farrell and Ioannou, this volume and Ioannou and Farrell, this volume), the initial perturbations of the ECMWF ensemble prediction system are based on the leading singular vectors of M (Buizza, this volume).

The relative performance of the singular-vector perturbations can be judged from Figure 1.4 (Buizza *et al.*, 2003), based on a comparison of ensemble prediction systems at ECMWF (Palmer *et al.*, 1993; Molteni *et al.*, 1996), NCEP (US National

Centers for Environmental Prediction; Toth and Kalnay, 1993) and MSC (Meteorological Service of Canada; Houtekamer *et al.*, 1996); the latter systems based on bred vectors and ensemble data assimilation respectively. The solid lines show the ensemble-mean root-mean-square error of each of the three forecast systems, the dashed lines show the spread of the ensembles about the ensemble mean. At initial time, both NCEP and MSC perturbations are inflated in order that the spread and skill are well calibrated in the medium range. The growth of perturbations in the ECMWF system, by contrast, appears to be more realistic, and overall the system appears better calibrated to the mean error.

1.3.2 Model uncertainty

From the discussion in the last section, part of the reason initial conditions are uncertain is that (e.g. in variational data assimilation) there is no rigorous operational procedure for comparing a model state X with an observation Y. The reason that there is no rigorous procedure is directly related to the fact that the model cannot be guaranteed to resolve well the circulation or weather features that influence the observation. In this respect model error is itself a component of initial error. Of course, model error plays an additional role as one integrates, forward in time, the model equations from the given initial state.

Unfortunately, there is no underlying theory which allows us to estimate the statistical uncertainty in the numerical approximations we make when attempting to integrate the equations of climate on a computer. Moreover, an assessment of uncertainty has not, so far, been a requirement in the development of subgrid parametrisations.

Parametrisation is a procedure to approximate the effects of unresolved processes on the resolved scales. The basis of parametrisation, at least in its conventional form, requires us to imagine that within a grid box there exists an ensemble of incoherent subgrid processes in secular equilibrium with the resolved flow, and whose effect on the resolved flow is given by a deterministic formula representing the mean (or bulk) impact of this ensemble. Hence a parametrisation of convection is based on the notion of the bulk effect of an incoherent ensemble of convective plumes within the grid box, adjusting the resolved scales back towards convective neutrality; a parametrisation of orographic gravity-wave drag is based on the notion of the bulk effect of an incoherent ensemble of breaking orographic gravity waves applying a retarding force to the resolved scale flow.

A schematic representation of parametrisation in a conventional weather or climate prediction model is shown in the top half of Figure 1.5. Within this framework, uncertainties in model formulation can be represented in the following hierarchical form:

- the multimodel ensemble whose elements comprise different weather or climate prediction models;

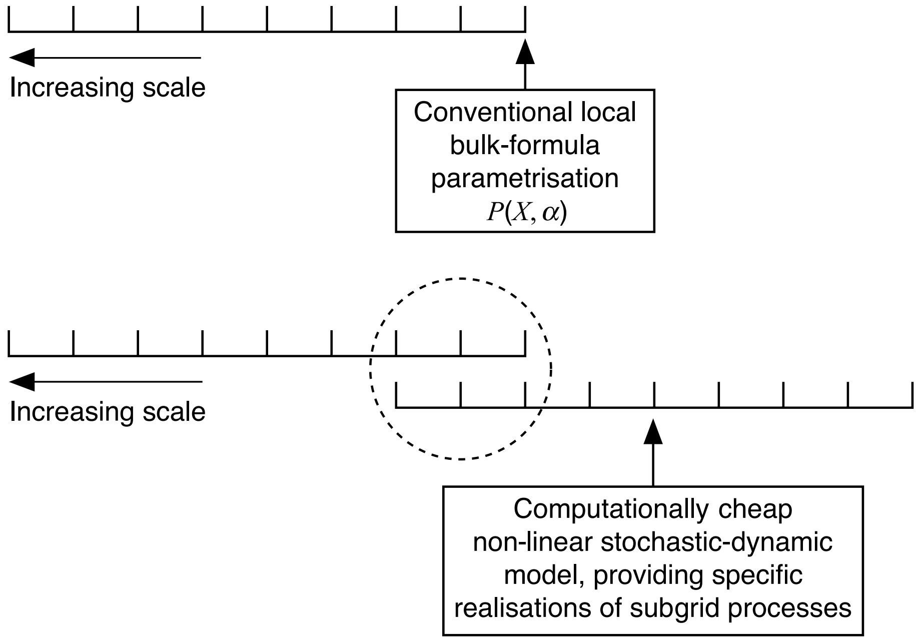

Figure 1.5 (Top) Schematic for conventional weather and climate prediction models (the 'Reynolds/Richardson Paradigm'). (Bottom) Schematic for weather and climate prediction using simplified stochastic-dynamic model representations of unresolved processes.

- the multiparametrisation ensemble whose elements comprise different parametrisation schemes $P(X, \alpha)$ within the same dynamical core;
- the multiparameter ensemble whose elements are all based on the same weather or climate prediction model, but with perturbations to the parameters α of the parametrisation schemes.

The DEMETER system (Palmer *et al.*, 2004; Hagedorn *et al.*, this volume) is an example of the multimodel ensemble; the ensemble prediction system of the Meteorological Service of Canada (Houtekamer *et al.*, 1996) is an example of a multiparametrisation scheme; the Met Office QUMP system (Murphy *et al.*, 2004) and the climateprediction.net ensemble system (Stainforth *et al.*, 2005) are examples of multiparameter ensemble systems.

The hierarchical representation of model error as discussed above should be considered a pragmatic approach to the problem – it certainly should not be considered a complete solution to the problem. The fundamental reason why parametrisations are uncertain is that in reality there is no scale separation between resolved and unresolved motions: according to Nastrom and Gage (1985), the observed spectrum of atmospheric motions shallows from a -3 slope to a $-5/3$ slope as the truncation limit of weather and climate models is approached. That is to say, the spectrum of

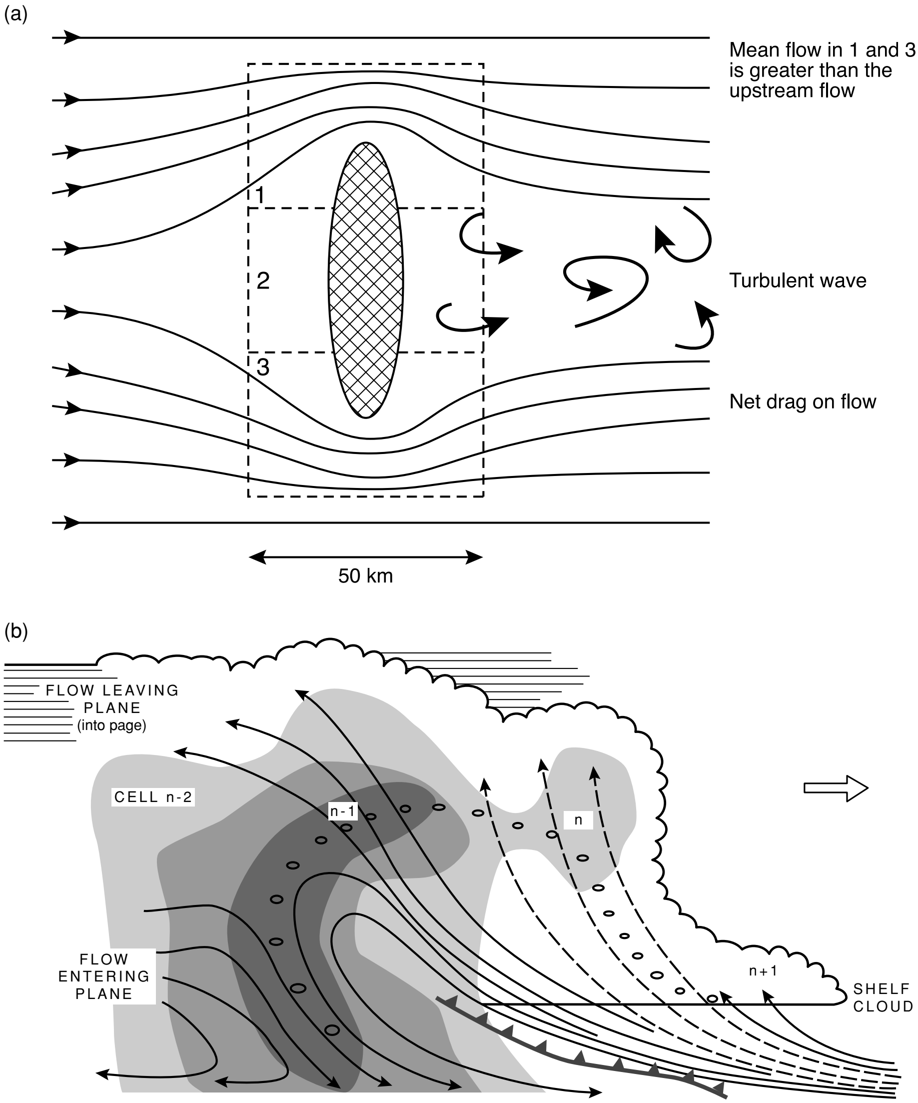

Figure 1.6 Schematic examples of the failure of conventional parametrisation to account for tendencies associated with subgrid processes: (a) when the subgrid topographic forcing is coherent across grid boxes; (b) when the convective motions have mesoscale organisation.

unresolved motions is dominated by a range of near grid-scale motions. Figure 1.6 gives schematic examples of near-grid-scale motions (a) for orographic flow and (b) for organised deep convection. In the case of orography, the flow is forced around the orographic obstacle. In the grid box containing the tip of the orography, a parametrisation will detect unresolved orography and apply a drag force, the very opposite of what may actually be required. In the case of convection, the grid box containing the bulk of the updraught may not be warming (through environmental subsidence); moreover, there is no requirement for the vertical momentum transfer to be upgradient, or for the implied kinetic energy generated by convective available potential energy within a grid box to be dissipated within that grid box.

Hence, in neither case can the impact of the unresolved orographic or convective processes be represented by conventional parametrisations, no matter what formulae are used or what values the underlying parameters take. In other words, part of the uncertainty in the representation of unresolved scales is structural uncertainty rather than parametric uncertainty.

In order to represent such structural uncertainty in ensemble prediction systems, we need to broaden the standard paradigmatic representation of subgrid scales. In considering this generalisation, it can be noted that in the conventional approach to weather and climate modelling, there is in fact a double counting of subgrid processes within an ensemble prediction system. By averaging across an ensemble prediction system with identical resolved-scale flow but different subgrid circulations, the ensemble prediction system effectively provides us with a mean of subgrid processes. But this averaging process has already been done by the parametrisation scheme, which itself is defined to be a mean of putative subgrid processes. Apart from possible inefficiency, what danger is there in such double counting?

The danger is that we miss a key element of the interaction between the resolved flow and the unresolved flow, leading to a component of systematic error in the climate models. Let us represent the grid-box tendency associated with unresolved scales by a probability density function $\rho_m(X)$ where X is some resolved-scale variable. Consider an ensemble prediction system where the grid-box mean variable is equal to X_0 across all members of the ensemble. Suppose now that instead of using the deterministic subgrid parametrisation $P(X_0;\alpha)$ across all ensemble members, we force the ensemble prediction system by randomly sampling $\rho_m(X_0)$. Would the ensemble-mean evolve differently? Yes, because of non-linearity!

It is therefore being suggested here that we change our paradigm of parametrisation as a deterministic bulk formula representing an ensemble of some putative 'soup' of incoherent subgrid processes, to one where each ensemble member is equipped with a possible realisation of a subgrid process. Hence, Figure 1.6(b) illustrates a possible generalisation in which the subgrid scales are represented by a simplified computationally efficient stochastic-dynamic system, potentially coupled to the resolved scales over a range of scales.

There are three reasons why the representation of model uncertainty through stochastic-dynamic parametrisation may be superior to the multimodel and related representations. Firstly, model uncertainty may be more accurately represented, and the corresponding ensembles may be more reliable. Secondly, as discussed above, noise-induced drift may lead to a reduction in model systematic error in a way impossible in a multimodel ensemble. Thirdly, estimates of 'natural climate variability' may be more accurate in a model with stochastic-dynamic representation of unresolved scales. This is important for the problem of detecting anthropogenic climate change.

A simple example of stochastic parametrisation has been discussed in Buizza *et al.* (1999) and Palmer (2001). Let us write, schematically, the equations of motion of our climate or weather prediction model as

$$\begin{aligned} \dot{X} &= F[X] + P + e \\ e &= \varepsilon P \end{aligned} \tag{1.9}$$

where P denotes the conventional parametrisation term and ε is a non-dimensional stochastic parameter with mean zero. The physical basis for such a multiplicative-noise form of stochastic parametrisation is that stochastic model perturbations are likely to be largest when the parametrisation tendencies themselves are largest, e.g. associated with intense convective activity, when the individual convective cells have some organised mesoscale structure, and therefore where the parametrisation concept breaks down. This notion of multiplicative noise has been validated from a coarse-grained budget analysis in a cloud-resolving model (Shutts and Palmer, 2004, 2006; Palmer *et al.*, 2005). Buizza *et al.* (1999) showed that probabilistic skill scores for the medium-range ensemble prediction systems (EPS) were improved using this stochastic parametrisation scheme.

From a mathematical point of view the addition of stochastic noise in Eq. (1.9) is straightforward. However, adding the noise term e makes a crucial conceptual difference to Eq. (1.9). Specifically, without stochastic noise, the subgrid parametrisation represents an averaged tendency associated with a supposed ensemble of subgrid processes occurring inside the grid box. With stochastic noise, the subgrid parametrisation represents a possible realisation of the subgrid tendency.

We can go further than this and ask whether some dynamical meteorology could be built into this stochastic realisation of the subgrid world. A possible stochastic-dynamic model could be associated with the cellular automaton, a computationally simple non-linear dynamical system introduced by the mathematician John von Neumann (Wolfram, 2002). Figure 1.7 (from Palmer, 2001) is a snapshot from a cellular automaton model where cells are either convectively active ('on') or convectively inhibited ('off'). The probability of a cell being 'on' is dependent on the convective available potential energy in the grid box and on the number of adjacent 'on' cells. Agglomerations of 'on' cells have the potential to feed vorticity from the parametrisation to the resolved scales. 'On' cells can be made to advect with the grid-mean wind. In a further development of this scheme (J. Berner, personal communication) a two-level multiscale cellular automaton has been developed. The smallest level represents individual convective plumes, whilst the intermediate level represents convectively coupled wave motions which can force the Madden–Julian Oscillation. It is planned to try to include characteristics of the dispersion equation of convectively coupled Kelvin waves in the intermediate cellular automaton.

Independently, cellular automata based on the Ising model have been developed as a stochastic-dynamic parametrisation of deep convection (Khouider *et al.*, 2003). Recently Shutts (2005) has built a hybrid stochastic-dynamic parametrisation scheme which combines the cellular automaton with the notion of stochastic

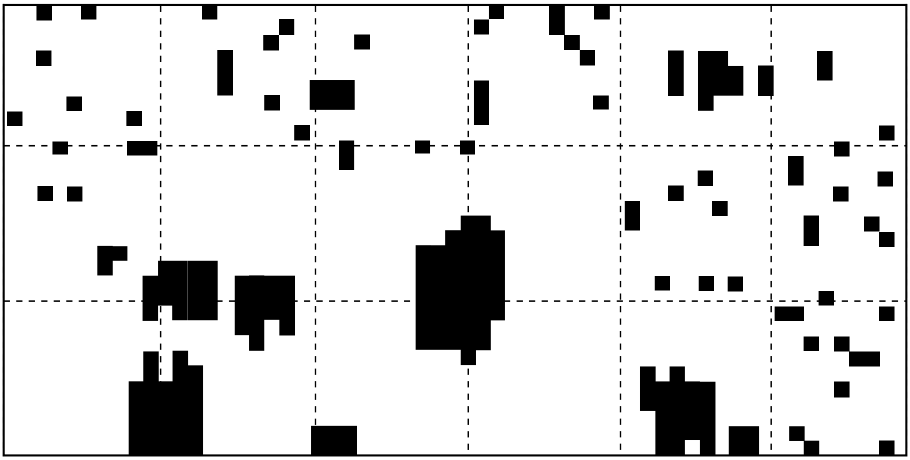

Figure 1.7 A snapshot in time from a cellular automaton model for convection. Black squares correspond to convectively active sites, white squares to convectively inhibited sites. From Palmer (2001).

backscatter (e.g. Leith, 1990; Frederiksen and Davies, 1997). Specifically, estimates of dissipation are made based on four components of the ECMWF model: convection, orographic gravity-wave drag, numerical diffusion, and the implicit dissipation associated with the semi-lagrangian scheme. The basic assumption is that a fraction (e.g. 10%) of the implied energy dissipation is actually fed back on to the model grid, through the cellular automaton.

Preliminary results with this scheme are very promising: there is quantifiable reduction in midlatitude systematic error which can be interpreted in terms of an increased frequency of occurrence of basic circulation regimes (Molteni *et al.*, this volume; Jung *et al.*, 2005) which are underestimated in the version of the model with conventional parametrisation. In this way, one can say that the stochastic parametrisation has reduced systematic error through a non-linear noise-induced drift effect.

1.4 Ensemble forecasts: some examples

In this section, some examples of ensemble forecasts from different forecast timescales will be shown, illustrating some of the ideas discussed above. Logically, one should perhaps order these examples by timescale, e.g. starting with the short range, finishing with centennial climate change. However, in this section the examples will be presented more or less in the historical order in which they were developed.

1.4.1 Extended range

Much of the early work on ensemble forecasting with comprehensive weather prediction models arose in trying to develop dynamical (as opposed to statistical-empirical)

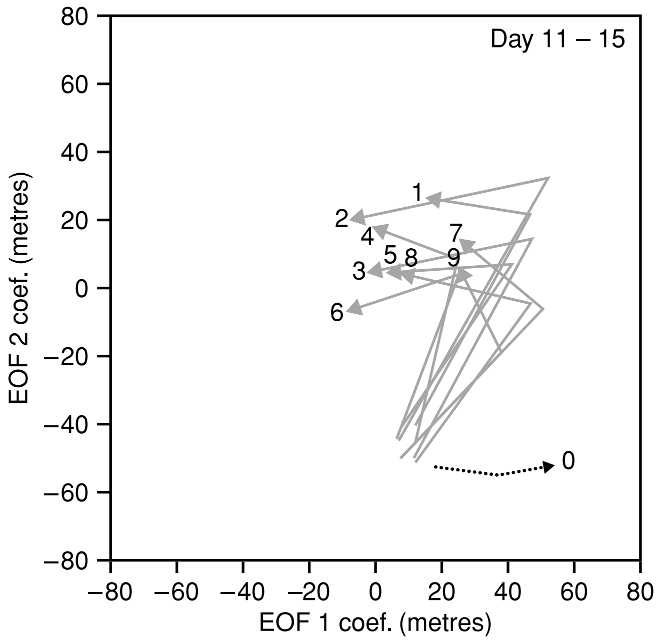

Figure 1.8 State-space trajectories from an unreliable time-lagged ensemble (solid), and verifying analysis (dashed). See Palmer *et al.* (1990).

techniques for monthly forecasting. One of the basic motivations for such work was that whilst the monthly timescale was clearly beyond the mean limit $[P_W]$ of predictability of daily weather, as shown by Miyakoda *et al.* (1986), from time to time the actual predictability of the atmospheric circulation would far exceed $[P_W]$. Ensemble forecasts were seen as a necessary means of determining ahead of time whether such enhanced predictability existed. The first ever operational probabilistic ensemble forecast was made for the 30-day timescale (Murphy and Palmer, 1986).

In these early days, methodologies to produce initial ensemble perturbations were rather simple, e.g. based on adding random noise to the initial conditions, or using the time-lagged technique, where ensemble members were taken from consecutive analyses (Hoffman and Kalnay, 1983).

Unfortunately, early results did not always live up to hopes. Figure 1.8 shows one of the pitfalls of ensemble forecasting. The figure shows the evolution in phase space (spanned by leading empirical orthogonal functions of the forecast ensemble) of an ensemble of five-day mean forecasts made using the time-lagged technique, based on an early version of the ECMWF model (Palmer *et al.*, 1990). Unfortunately the evolution of the verifying analysis is in a class of its own! A probability forecast based on this ensemble would clearly be unreliable. Charlie would not be impressed!

During the 1990s, work on 30-day forecasting went into a period of decline. However, in 2004, an operational 30-day forecast system was finally implemented at ECMWF, using singular-vector ensemble perturbations in a coupled ocean–atmosphere model (Vitart, 2004). Unlike the early example shown above, probability forecasts have been shown to be reliable in the extended range. This improvement results from developments in data assimilation, in deterministic forecasting, and in medium-range ensemble forecasting as discussed in this chapter and in Buizza (this volume).

1.4.2 Medium range

A general assessment of predictability in the medium range, based on the ECMWF system, is given in Simmons (this volume). Even though early results on 30-day forecasting were disappointing, it was nevertheless clear that the idea of using ensemble forecasts to determine periods where the atmospheric circulation was either especially predictable, or especially unpredictable, was also relevant to the medium range. Based on the experience outlined above, there was clearly a need to ensure that the resulting ensembles were not underdispersive. The initial work in this area was done at ECMWF and NCEP, using different methods for obtaining initial perturbations (see section above). The ECMWF ensemble prediction system comprises 51 forecasts using both singular vector initial perturbations and stochastic physics (more details are given in Buizza, this volume).

In late December 1999, two intense storms, subsequently named Lothar and Martin, ran across continental Europe leaving behind a trail of destruction and misery, with over 100 fatalities, over 400 million trees blown down, over 3 million homes without electricity and water. Figure 1.9 shows the ensemble 'stamp maps' (based on a TL255 version of the ECMWF model) for Lothar, at initialisation time on 24 December and for forecast time 6 UTC on 26 December. This storm was exceptionally unpredictable, and even at 42 hours lead time there is considerable spread in the ensemble. The best-guidance deterministic forecast only predicts a weak trough in surface pressure. A number of members of the ensemble support this forecast; however, a minority of ensemble members also show an intense vortex over France. In this sense, the ensemble was able to predict the risk of a severe event, even though it was impossible to give a precise deterministic forecast. More recent deterministic reforecasts with a T799 version of the ECMWF model also fail to predict this storm (Martin Miller, personal communication) – this is clearly a case which demonstrates the value of ensemble forecasts even at intermediate resolution (Palmer, 2002).

1.4.3 Seasonal and decadal prediction

The scientific basis for seasonal prediction lies in the interaction of the atmosphere with slowly varying components of the climate system: notably the oceans and land surface (Timmermann, this volume; Shukla and Kinter, this volume). Early work showed firstly that El Niño events are predictable seasons ahead of time using intermediate-complexity coupled ocean–atmosphere models of the tropical Pacific

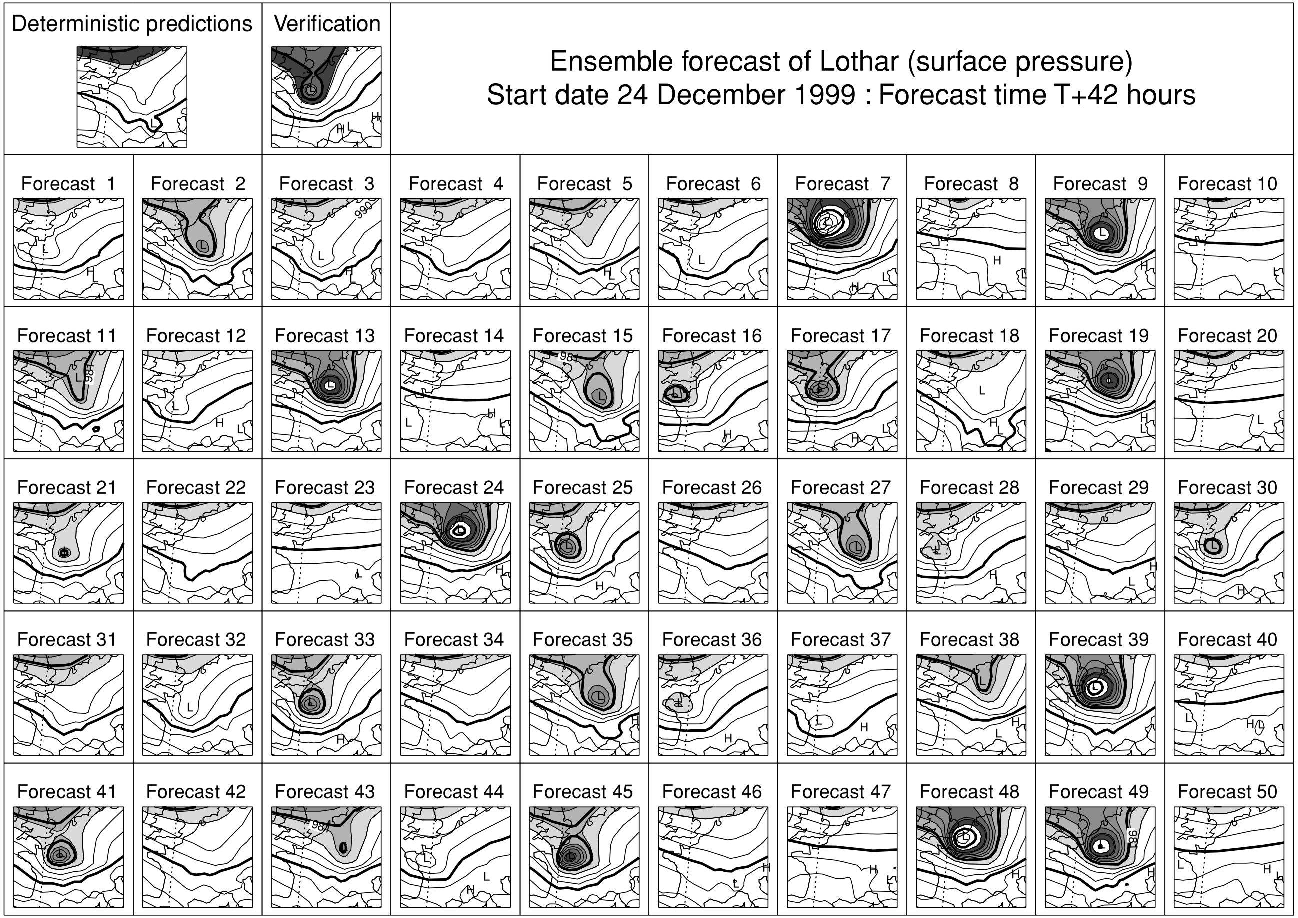

Figure 1.9 Isopleths of surface pressure from a 51-member ECMWF ensemble forecast of the exceptionally severe storm 'Lothar', based on initial conditions 42 hours before the storm crossed northern France in December 1999. The top left shows the forecast made from the best estimate of these initial conditions. This did not indicate the presence of a severe storm. The 50 forecasts which comprise the ensemble indicated exceptional unpredictability and a significant risk of an intense vortex.

(Zebiak and Cane, 1987), and secondly that sea surface temperature (SST) anomalies in the tropical Pacific Ocean have a global impact on atmospheric circulation (e.g. Shukla and Wallace, 1983). Putting these factors together has led to global seasonal prediction systems based on comprehensive global coupled ocean–atmosphere models (Stockdale *et al.*, 1998). Inevitably such predictions have been based on ensemble forecast techniques, where initial perturbations represent uncertainties in both atmosphere and ocean analyses.

In addition to initial uncertainty, representing forecast model uncertainty is a key element in reliably predicting climate risk on seasonal and longer timescales. The ability of multimodel ensembles to produce more reliable forecasts of seasonal climate risk over single-model ensembles was first studied by the PROVOST (Prediction of Climate Variations on Seasonal to Interannual Timescales) project funded by the European Union IVth Framework Environment Programme, and a similar 'sister' project DSP (Dynamical Seasonal Prediction) undertaken in the United States (Palmer and Shukla, 2000).

As part of the PROVOST project, three different atmospheric general circulation models (including one model at two different resolutions) were integrated over four-month timescales with prescribed observed SSTs. Each model was run in ensemble mode, based on nine different initial conditions from each start date; results were stored in a common archive. One of the key results from PROVOST and DSP was that, despite identical SSTs, ensembles showed considerable model-to-model variability in estimates both of the SST-forced seasonal-mean signal, and the seasonal-mean 'noise' generated by internal dynamics (Straus and Shukla, 2000). Consistent with this, probability scores based on the full multimodel ensemble scored better overall than any of the individual model ensembles (e.g. Doblas-Reyes *et al.*, 2000; Palmer *et al.*, 2000).

Based on such results, the DEMETER project (Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction; Palmer *et al.*, 2004, Hagedorn *et al.*, this volume) was conceived, and successfully funded under the European Union Vth Framework Environment Programme. The principal aim of DEMETER was to advance the concept of multimodel ensemble prediction by installing a number of state-of-the-art global coupled ocean–atmosphere models on a single supercomputer, and to produce a series of six-month ensemble reforecasts with common archiving and common diagnostic software.

Figure 1.10 shows an example of results from DEMETER. Forecasts of El Niño are seen to be more reliable in the multimodel ensemble than in the ECMWF single-model ensemble; more specifically, the observed SSTs do not always lie in the range of the ECMWF-model ensembles, but do lie in the range of the DEMETER multimodel ensembles. Other results supporting the notion that multimodel ensembles are more reliable than single-model ensembles are given in Hagedorn *et al.* (this volume). However, it is not necessarily the case that multimodel ensembles are reliable for all variables. As discussed in Palmer *et al.* (2005), seasonal forecasts of

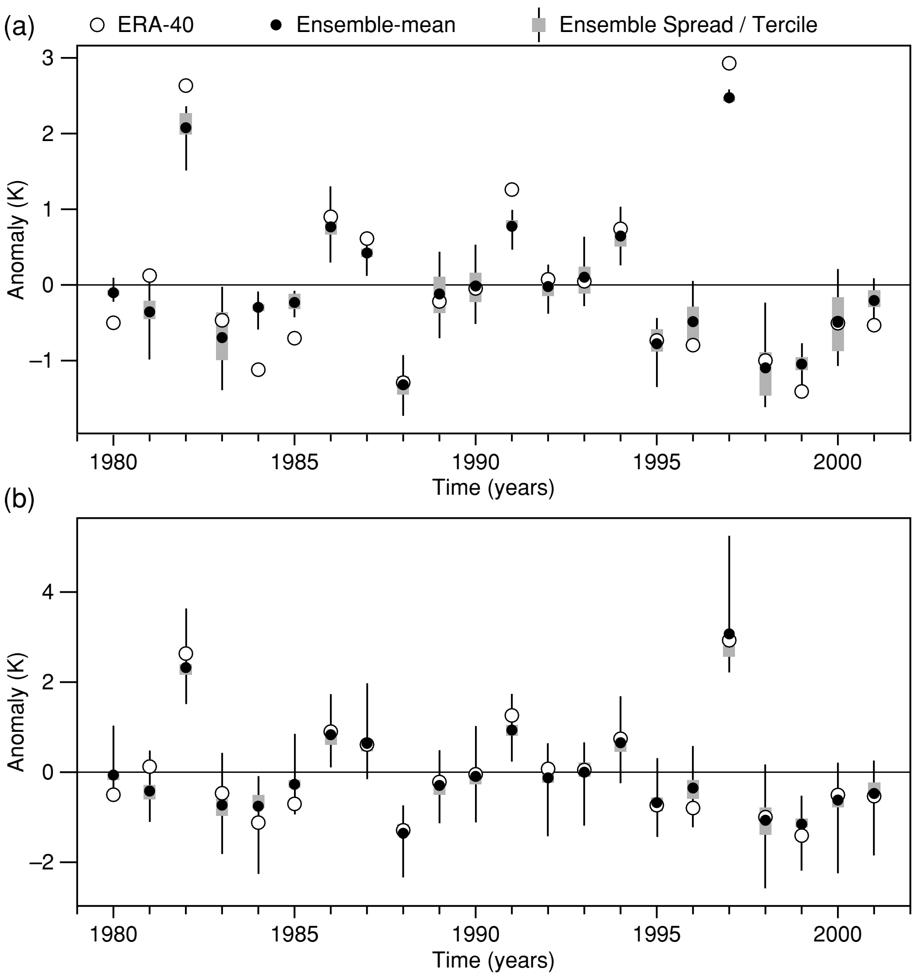

Figure 1.10 Time series of forecast NINO-3 SST anomalies for DJF, initialised 1 November, based on (a) ECMWF ensemble, (b) DEMETER multimodel ensemble. Bars and whiskers show terciles, the ensemble-mean values are shown as solid circles, and the actual SST anomalies are shown as open circles.

upper-tercile precipitation over Europe are neither reliable in single-model nor DEMETER multimodel ensemble systems. Consistent with the discussion above, this latter result suggests that multimodel systems may not represent model uncertainty completely. In the European Union FP6 project ENSEMBLES, it is proposed to compare the reliability of seasonal ensemble forecasts, made using the multimodel technique, with stochastic parametrisation.

At the beginning of this chapter, the existence or otherwise of predictability was discussed from the perspective of decision-making: are the forecast probability densities sufficiently different from climatological densities to influence decision-making? In DEMETER quantitative crop and malaria prediction models were linked to individual ensemble members; based on this, the probability of crop failure or malaria epidemic could be estimated. See Hagedorn *et al.* (this volume) for details.

Many observational and modelling studies document pronounced decadal and multidecadal variability in the Atlantic, Pacific and Southern Oceans. For example, decadal variability in Atlantic sea surface temperatures is in part associated with fluctuations in the thermohaline circulation (e.g. Broecker, 1995); decadal variability in the Pacific is associated in part with fluctuations in the Pacific Decadal Oscillation (e.g. Barnett *et al.*, 1999). Variations in the Atlantic thermohaline circulation appear to be predictable one or two decades ahead, as shown by a number of perfect model predictability studies, e.g. Griffies and Bryan (1997), Latif *et al.* (this volume). The SST anomalies (both tropical and extratropical) associated with decadal variations in the thermohaline circulation appear to impact the North Atlantic Oscillation in the extratropics. A weakening of the thermohaline circulation is likely to lead to a cooling of northern European surface temperatures. There is also evidence (Landerer *et al.*, 2006) that decadal variability in the thermohaline circulation could lead to significant decadal sea level fluctuations in Europe.

Variations in the thermohaline circulation may also be associated with climate fluctuations in the tropics. For example, there is some evidence that long-lasting drought over the African Sahel is associated with decadal-timescale variability in the so-called sea surface temperature dipole in the tropical Atlantic (Folland *et al.*, 1986).

In order for decadal prediction to evolve into a possible operational activity, suitable observations from which the thermohaline circulation can be initialised must exist. Programmes such as ARGO (Wilson, 2000) may help provide such observations. These need to be properly assimilated in the ocean component of a coupled forecast model. It is clearly essential that the model itself has a realistic representation of the thermohaline circulation.

On the timescale of a decade, anticipated changes in greenhouse gas concentrations will also influence the predictions of future climate (Smith *et al.*, 2006). In this sense, decadal timescale prediction combines the pure initial value problem with the forced climate problem, discussed in the next subsection.

1.4.4 Climate change

Climate change has been described by the UK Government's Chief Scientific Advisor as one of the most serious threats facing humanity – more serious even than the terrorist threat. Nevertheless, there is uncertainty in the magnitude of climate change; this uncertainty can be quantified using ensemble techniques (Allen *et al.*, this volume). For example, Palmer and Räisänen (2002) used the multimodel ensemble technique to assess the impact of increasing levels of CO_2 on the changing risk of extreme seasonal rainfall over Europe in winter (Figure 1.11; colour plate), and also for the Asian summer monsoon, based on the CMIP multimodel ensemble

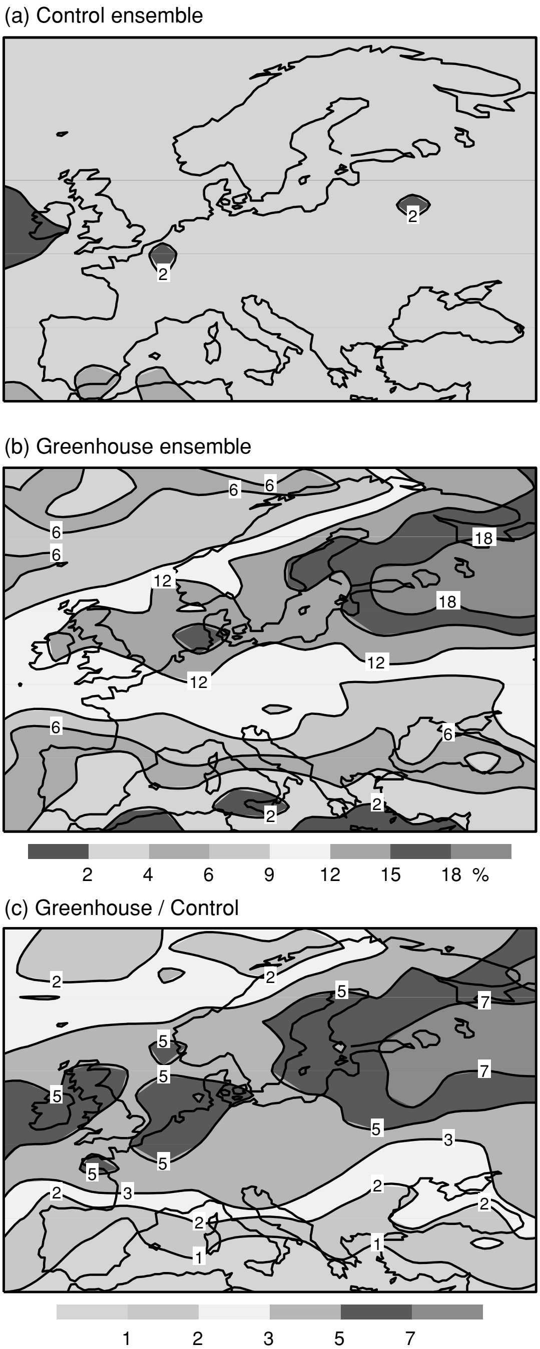

Figure 1.11 (See also colour plate section.) The changing probability of extreme seasonal precipitation for Europe in boreal winter. (a) The probability (in %) of a 'very wet' winter defined from the control CMIP2 multimodel ensemble with twentieth-century levels of CO_2 and based on the event E: total boreal winter precipitation greater than the mean plus two standard deviations. (b) The probability of E but using data from the CMIP multimodel ensemble with transient increase in CO_2 and calculated around the time of CO_2 doubling (years 61–80 from present). (c) The ratio of values in (b) to those in (a), giving the change in the risk of a 'very wet' winter arising from human impact on climate. From Palmer and Räisänen (2002).

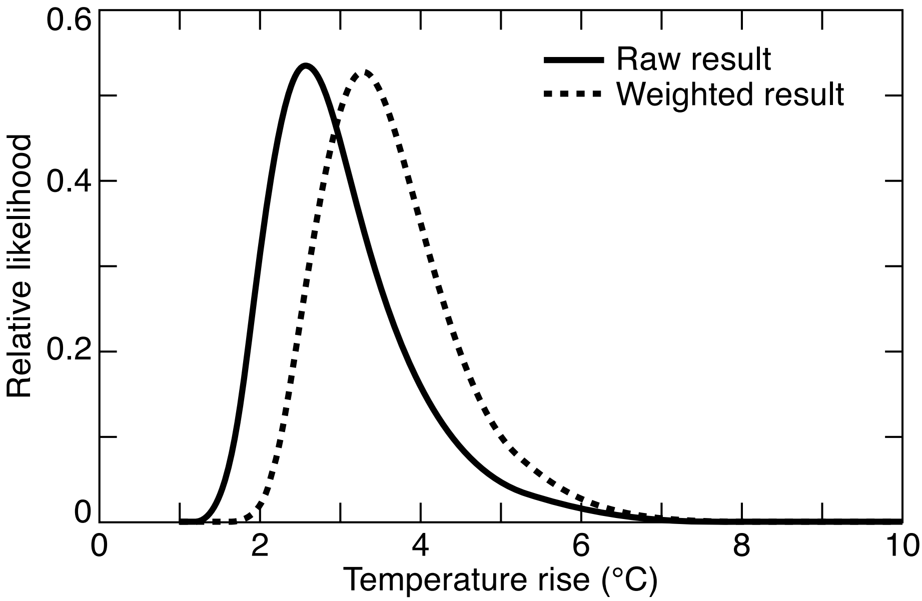

Figure 1.12 Probability distributions of global average annual warming associated with a 53-member ensemble for a doubling of carbon dioxide concentration. Ensemble members differ by values of key parameters in the bulk formulae used to represent unresolved processes, in a version of the Hadley Centre climate model. Solid curve: based on 'raw model output'. Dashed curve: the probability distribution weighted according to the ability of different model versions to simulate observed present day climate. From Murphy *et al.* (2004).

(Meehl *et al.*, 2000). More recently Murphy *et al.* (2004) and Stainforth *et al.* (2005) have quantified uncertainty in climate sensitivity (the global warming associated with a doubling of CO_2) based on multiparameter ensembles (see also Allen *et al.*, this volume).

Figure 1.12 shows a probability distribution of climate sensitivity based on the multiparameter ensemble of Murphy *et al.* (2004). The solid line shows the raw output; the dashed line shows results when the individual ensemble members are weighted according to the fit of control integrations to observations. It is interesting to note that this fit to data does not change the range of uncertainty – rather the forecast probability distribution is shifted to larger values of climate sensitivity.

As yet, probability distributions in global warming have not been estimated using the stochastic physics approach. It will be interesting to see how estimates of uncertainty in climate sensitivity are influenced by the methodology used to represent model uncertainty.

How can the uncertainty in global warming be reduced? It can be noted that much of the ensemble spread in Figure 1.12 is associated with uncertainty in parameters from parametrisations of clouds and boundary layer processes. These are fast-timescale processes. Hence it may be possible to reduce the spread in Figure 1.12 by assessing how well the contributing models perform as short-range weather prediction models. More specifically, the type of budget residual technique pioneered by Klinker and Sardeshmukh (1992) could be applied to the multiparameter ensemble. On this basis, models with large residuals, obtained by integrating over just a few time steps but from many different initial conditions, could be rejected from the ensemble (Rodwell and Palmer, 2006).

Ultimately, the uncertainties in climate prediction arise because we are not solving the full partial differential equations of climate – for example, the important cloud processes mentioned above are parametrised, and, as noted, parametrisation theory is rarely a justifiable procedure. On this basis, further reduction of the uncertainty in global warming may require significantly larger computers so that at least major convective cloud systems can be resolved. Of course, this will not eliminate uncertainty, as cloud microphysics will still have to be parametrised.

In truth, reducing uncertainty in forecasts of climate change will require a combination of significantly greater computer resources and the use of sophisticated validation techniques as used in numerical weather prediction studies.

1.4.5 Short-range forecasting

For many years, it was generally assumed that while ensemble techniques may well be important for medium and longer range predictions, the short-range weather prediction problem, up to day 2, let's say, should essentially be considered deterministic. Such a view is no longer held today – predictions of flash floods and other types of mesoscale variability are not likely to be strongly predictable on timescales of a day. To quantify uncertainty in such forecasts, ensemble prediction systems based on multiple integrations of fine-scale limited area models are now being actively developed around the world. To create such ensembles, ensemble boundary conditions are taken from a global ensemble prediction system, and these are combined with initial perturbations within the limited-area model domain. Insofar as some of the principal forecast variables are related to processes close to the model resolution, a representation of model uncertainty is also necessary.

An example of an ensemble of limited-area model integrations is shown in Figure 1.13, based on the COSMO-LEPS system (Tibaldi *et al.*, this volume; Waliser *et al.*, 2006). The boundary conditions for this limited-area model ensemble have been taken from ECMWF ensemble integrations (in this case using moist singular vectors; M. Leutbecher, personal communication). The example shown here is for the storm Lothar (see Figure 1.9). It can be seen that the limited-area model ensemble is predicting a significant risk of damaging wind gusts – in a situation where the deterministic forecast from the most likely initial state had no warning of severe gusts at all.

Development of short-range ensemble prediction systems using limited-area models is now a significant growth area.

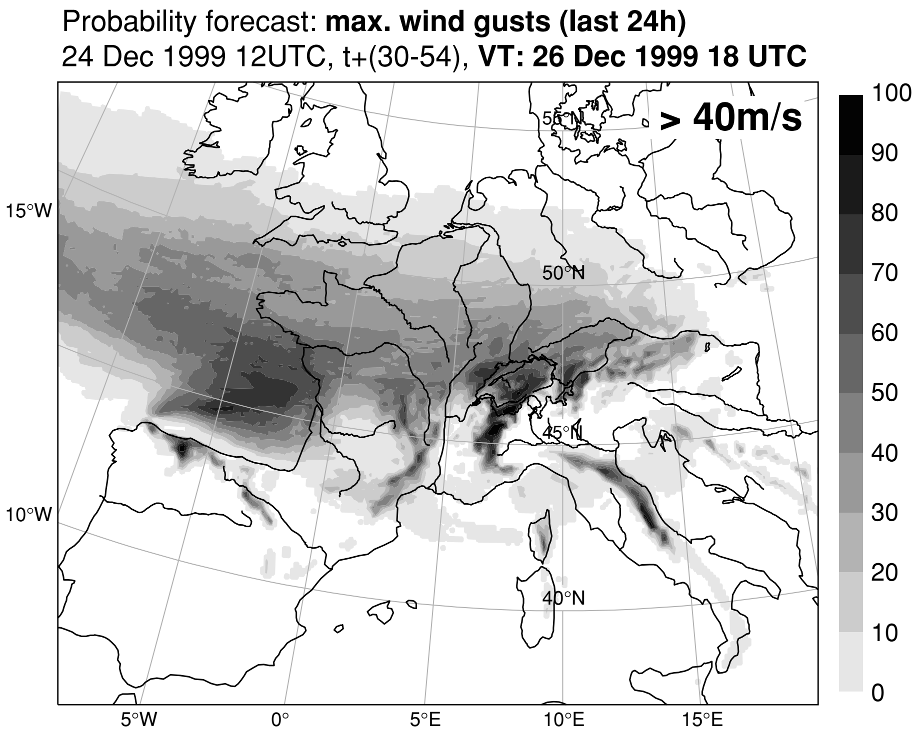

Figure 1.13 Probability that wind gusts exceed 40 m/s for the storm Lothar based on an ensemble (COSMO-LEPS) of limited-area model integrations using ECMWF ensemble boundary conditions. From Waliser *et al.* (2005).

1.5 Discussion

In this chapter, we have charted a revolution in the way weather and climate predictions are produced and disseminated as probability forecasts. The revolution started in studies of monthly predictability, spread to medium range and seasonal timescales, finally permeating extremes of meteorological predictions, the climate change and short-range weather forecast problems. The revolution is based on the notion that in many cases the most relevant information for some user is not necessarily what is most likely to happen, but rather a quantified probability of weather or climate events to which the user is sensitive. We introduced the case of Charlie, who wants to know if he can lay concrete. If Charlie's cost/loss ratio is less than 0.5, he may decide not to lay concrete even when frost is not likely to occur. If Charlie were one day to become minister for the environment, he might be faced with similar risk-based decisions on the adequacy of the current Thames barrier, one of London's key flood defences. A prediction of an 11 K warming associated with a doubling of CO_2 (Stainforth *et al.*, 2005) doesn't have to be likely in order for the replacement of the barrier to be in need of urgent consideration. However, the decision to replace may need better quantified estimates of uncertainty than we currently have.

Predicting the probability of occurrence of weather and climate events requires us to be able to quantify the sources of uncertainty in weather and climate prediction, and to estimate how these sources actually impact on the predictions themselves. In practice, these sources of uncertainty are not easy to quantify. This is not because we don't know the accuracy of instruments used to observe the atmosphere (and oceans). Rather, it is because the approximations used in making computational models of the essentially continuum multiphase fluid equations are themselves hard to quantify. Hence, for example, when model variables are compared with observations in data assimilation, the observation operator doesn't recognise the fact that the observation may be strongly influenced by scales of motion that the model is unable to resolve well. On many occasions and for certain types of observation (e.g. surface pressure) this may not be a serious problem, but occasionally and for other types of observation (e.g. humidity) it is. At present, this type of uncertainty is unquantified in operational data assimilation – from this perspective it is an example of an 'unknown unknown'. In the presence of such 'unknown unknowns', operational ensemble prediction systems run the danger of being underdispersive. This is potentially disastrous: if Charlie lays concrete when the risk of frost is predicted to be zero, and frost occurs, Charlie will never use ensemble prediction again! If, in his future career as politician, Charlie decides against replacing the Thames barrier on the basis of underdispersive ensemble climate forecasts, history may not be kind to him!

One specific conclusion of this chapter is that the development of accurate ensemble prediction systems on all timescales, hours to centuries, relies on a better quantification of model uncertainties. It has been argued that this may require a fundamental change in the way we formulate our models, from deterministic to stochastic dynamic. This change has been anticipated by Lorenz (1975) who said: 'I believe that the ultimate climatic models . . . will be stochastic, i.e. random numbers will appear somewhere in the time derivatives'. Stochastic representations of subgrid processes are particularly well suited to ensemble forecasting.

References

Barnett, T. P., D. W. Pierce, M. Latif, D. Dommenget and R. Saravanan (1999). Interdecadal interactions between the tropics and midlatitudes in the Pacific basin. *Geophys. Res. Lett.*, **26**, 615–18.

Broecker, W. (1995). Chaotic climate. *Sci. Am.*, November, 62–8.

Buizza, R. and T. N. Palmer (1995). The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–56.

Buizza, R., M. J. Miller and T. N. Palmer (1999). Stochastic simulation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–908.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei and Y. Zhu (2003). Assessment of the status of global ensemble prediction. *Proceedings of 2002 ECMWF Seminar on Predictability of Weather and Climate*. ECMWF.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei and Y. Zhu (2005). A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–97.

Courtier, P., E. Andersson, W. Heckley, *et al.* (1998). The ECMWF implementation of three dimensional variational assimilation (3DVAR). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–808.

Doblas-Reyes, F. J., M. Deque and J.-P. Piedlievre (2000). Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.*, **126**, 2069–88.

Folland, C. K., T. N. Palmer and D. E. Parker (1986). Sahel rainfall and worldwide sea surface temperatures 1901–85. *Nature*, **320**, 602–7.

Frederiksen, J. S. and A. G. Davies (1997). Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.*, **54**, 2475–92.

Griffies, S. M. and K. Bryan (1997). A predictability study of simulated North Atlantic multidecadal variability. *Clim. Dynam.*, **13**, 459–87.

Hoffman, R. and E. Kalnay (1983). Lagged average forecasting, alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–18.

Houtekamer, P., L. Lefaivre, J. Derome, H. Ritchie and H. Mitchell (1996). A system simulation approach to ensemble prediction. *Mon. Weather Rev.*, **124**, 1225–42.

Jung, T., T. N. Palmer and G. J. Shutts (2005). Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.*, **32**, L23811.

Klinker, E. and P. D. Sardeshmukh (1992). The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–27.

Khouider, B., A. J., Majda and M. A. Katsoulakis (2003). Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci. USA*, **100**, 11941–6.

Landerer, F., J. Jungclaus, J. Marotzke (2006). Steric and dynamic change in response to the A1B scenario integration in the ECHAM5/MPI-OM coupled climate model. *J. Phys. Oceanogr.*, in press.

Leith, C. (1990). Stochastic backscatter in a sub-gridscale model: plane shear mixing layer. *Phys. Fluids A*, **2**, 297–9.

Lewis, J. M. (2005). Roots of ensemble forecasting. *Mon. Weather Rev.*, **133**, 1865–85.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, **42**, 433–71.

(1965). A study of the predictability of the 28-variable atmospheric model. *Tellus*, **17**, 321–33.

(1975). Climatic predictability. In *The Physical Basis of Climate and Climate Modelling*. WMO GARP Publication Series No. 16. World Meteorological Organisation.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif and R. J. Stouffer (2000). The Coupled model intercomparison project. *Bull. Am. Meteorol. Soc.*, **81**, 313.

Miyakoda, K., J. Sirutis and J. Ploshay (1986). One month forecast experiments – without anomaly boundary forcings. *Mon. Weather Rev.*, **114**, 1145–76.

Molteni, F. and T. N. Palmer (1993). Predictability and non-modal finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–98.

Molteni, F., R. Buizza, T. Petroliagis and T. N. Palmer (1996). The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Murphy, A. H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Weather Rev.*, **105**, 803–16.

Murphy, J. M. and T. N. Palmer (1986). Experimental monthly long-range forecasts for the United Kingdom. II: A real time long-range forecast by an ensemble of numerical integrations. *Meteorol. Mag.*, **115**, 337–49.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, *et al.* (2004). Quantifying uncertainties in climate change using a large ensemble of global climate model predictions. *Nature*, **430**, 768–72.

Nastrom, G. D. and K. S. Gage (1985). A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–60.

Palmer, T. N. (1988). Medium and extended range predictability and stability of the Pacific/North American mode. *Quart. J. Roy. Meteor. Soc.*, **114**, 691–713.

(2001). A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrisation in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304.

(2002). The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–74.

Palmer, T. N. and J. Shukla (2000). Editorial to DSP/PROVOST special issue, *Quart. J. Roy. Meteor. Soc.*, **126**, 1989–90.

Palmer, T. N. and J. Räisänen (2002). Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, **415**, 512.

Palmer, T. N., C. Brankovic, F. Molteni, *et al.* (1990). The European Centre for Medium-Range Weather Forecasts (ECMWF) program on extended-range prediction. *Bull. Am. Meteorol. Soc.*, **71**, 1317–30.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet and J. Tribbia (1993). Ensemble prediction. In *ECMWF 1992 Seminar: Validation of Models over Europe*. ECMWF.

Palmer, T. N., C. Brankovic and D. S. Richardson (2000). A probability and decision-model analysis of PROVOST seasonal multimodel ensemble integrations. *Quart. J. Roy. Meteor. Soc.*, **126**, 2013–34.

Palmer, T. N., A. Alessandri, U. Andersen, *et al.* (2004). Development of a European multi-model ensemble system for seasonal to inter-annual prediction. *Bull. Am. Meteorol. Soc.*, **85**, 853–72.

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung and M. Leutbecher (2005). Representing model uncertainty in weather and climate prediction. *Ann. Rev. Earth Planet. Sci.*, **33**, 163–93.

Reynolds, C. and T. N. Palmer (1998). Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.*, **55**, 3005–23.

Rodwell, M. J. and T. N. Palmer (2006). Assessing model physics with initial forecast tendencies: application to climate change uncertainty. *Quart. J. Roy. Meteor. Soc.*, Submitted.

Shukla, J. and J. M. Wallace (1983). Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **40**, 1613–30.

Shutts, G. J. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–102.

Shutts, G. J. and T. N. Palmer (2004). The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In *Proceedings of ECMWF Workshop on Intra-seasonal Variability*. ECMWF.

(2006). Convective forcing fluctuations in a cloud-resolving model: relevance to the stochastic parametrization problem. *J. Clim.*, in press.

Smith, D. M., A. W. Colman, S. Cusack, C. K. Folland, S. Ineson and J. M. Murphy (2006). Predicting surface temperature for the coming decade using a global climate model. *Nature*, in press.

Stainforth, D. A., T. Aina, C. Christensen, *et al.* (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–6.

Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves and M. A. Balmaseda (1998). Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370–3.

Straus, D. M. and J. Shukla (2000). Distinguishing between the SST-forced variability and internal variability in mid-latitudes: analysis of observations and GCM simulations. *Quart. J. Roy. Meteor. Soc.*, **126**, 2323–50.

Toth, Z. and E. Kalnay (1993). Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317–30.

Vitart, F. (2004). Monthly forecasting at ECMWF. *Mon. Weather Rev.*, **132**, 2761–79.

Waliser, A., M. Arpagaus, M. Leutbecher and C. Appenzeller (2006). The impact of moist singular vectors and horizontal resolution on short-range limited-area ensemble forecasts for two European winter storms. *Mon. Weather Rev.*, in press.

Wilson, S. (2000). Launching the Argo armada. *Oceanus*, **42**, 17–19.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media Inc.

Zebiak, S. and M. Cane (1987). A model of the El Niño-Southern Oscillation. *Mon. Weather Rev.*, **115**, 2262–78.

2

Predictability from a dynamical meteorological perspective

Brian Hoskins
University of Reading

2.1 Introduction: origins of predictability

Predictability of weather at various timescales has its origins in the physics and dynamics of the system. The annual cycle is an example of very predictable behaviour with such an origin, though this predictability is of little other than general background use to the forecaster. The rapid rotation of the Earth with its shallow, generally stably stratified atmosphere leads to the dominance of phenomena with balance between their thermodynamic and dynamic structures. These structures evolve on timescales comparable to, or longer than, a day. This balanced motion is best described by consideration of two properties that are materially conserved under adiabatic conditions, potential temperature (θ, or equivalently entropy) and potential vorticity (PV). A feature of particular importance, leading to potentially predictable behaviour, is the ability of the atmosphere to support balanced, large-scale Rossby waves.

Atmospheric phenomena with recognised structures tend to exhibit characteristic evolutionary behaviour in time and thus to have a level of predictability. Particular examples of such phenomena are the midlatitude cyclone on synoptic timescales, the tropical Intra-Seasonal Oscillation and the El Niño–Southern Oscillation (ENSO) on annual timescales.

Slower parts of the climate system can leave an imprint on shorter timescales and hence give an element of predictability to them. Tropical sea surface temperature (SST) anomalies tend to persist and can lead to anomalous convective activity in

Predictability of Weather and Climate, ed. Tim Palmer and Renate Hagedorn. Published by Cambridge University Press.

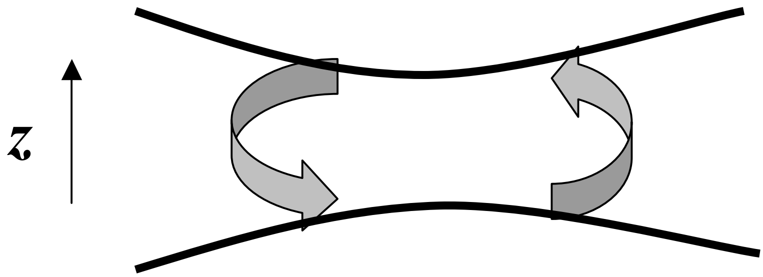

Figure 2.1 The balanced structure typically associated with a positive potential vorticity (PV) anomaly. Features shown are the cyclonic circulation in the region and the isentropes 'sucked' towards the anomaly in the vertical, consistent with high static stability there.

their region. This activity can trigger Rossby waves that communicate anomalous conditions to other regions of the globe. Soil moisture anomalies in tropical or extra-tropical regions can persist a month or more and similarly trigger both local and remote responses.

Here aspects of balanced motion and its description with PV and Rossby waves will be discussed in Section 2.2. The focus in Section 2.3 is on particular phenomena, midlatitude weather systems, blocking highs, and a particular mode of variability, the North Atlantic Oscillation (NAO). Section 2.4 gives some discussion of aspects of the summer of 2002 and their possible predictability. Some concluding comments are given in Section 2.5.

2.2 Balanced motion, potential vorticity and Rossby waves

Hydrostatic balance and geostrophic balance together lead to so-called thermal wind balance between the wind and temperature fields. The development of such motion is described by the quasi-geostrophic version of the conservation of PV. As discussed in detail in Hoskins *et al.* (1985), more general balanced motion is uniquely determined through an elliptic problem by the PV/θ distribution. Its development is described by the material conservation of PV on θ surfaces and θ on the lower boundary. Alternatively it is often convenient to summarise the upper tropospheric PV/θ distribution by the distribution of θ on the PV $=$ 2PVU surface (here northern hemisphere signs are used for convenience), which away from the tropics can be considered to be the dynamical tropopause (Hoskins, 1997).

Because of the elliptic nature of the inversion problem, a positive PV anomaly is generally associated with both cyclonic motion and increased static stability (Figure 2.1). Similarly a negative PV anomaly is associated with anticyclonic motion and reduced static stability. As in Figure 2.2, the tip of PV 'trough' often elongates,

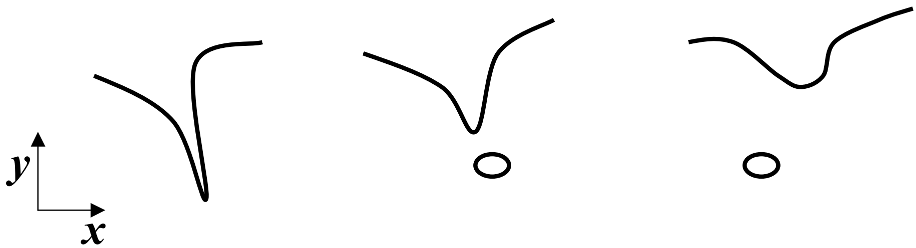

Figure 2.2 An upper air development of a geopotential trough, indicated by a PV contour on a θ surface or a θ contour on a PV surface in the tropopause region. Shown is the elongation of the trough, and the development and movement away of a cut-off.

develops its own cyclonic circulation and cuts itself off. Once it has done this, the cut-off low must continue to exist until either the PV anomaly is eroded by diabatic processes or it moves back into the higher PV region. This implies some extended predictability of such a cut-off low once it has formed.

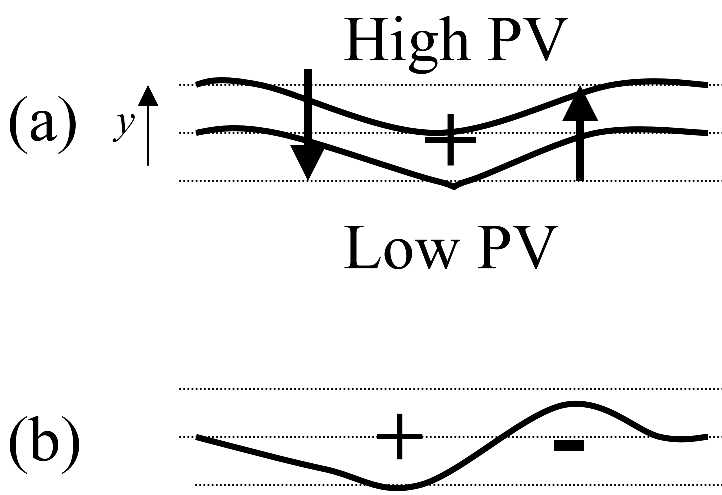

Figure 2.3 Rossby wave development. In (a) PV contours are shown displaced to the south, leading to a PV anomaly and cyclonic circulation. This circulation advects the PV meridionally, leading to the PV distribution and anomalies shown in (b).

If a contour is displaced equatorwards as in Figure 2.3(a), the associated cyclonic anomaly advects the PV distribution as shown in Figure 2.3(b). This gives a negative (anticyclonic) PV anomaly to the east and a positive (cyclonic) anomaly to the west. If there is a basic westerly flow, the net result is that the original PV anomaly will move to the east at less than the speed of the basic flow and could be stationary. However, the wave activity develops downstream, to the east, at a speed greater than that of the basic flow. The motion described is that of Rossby waves and the two speeds described respectively their phase and group speeds.

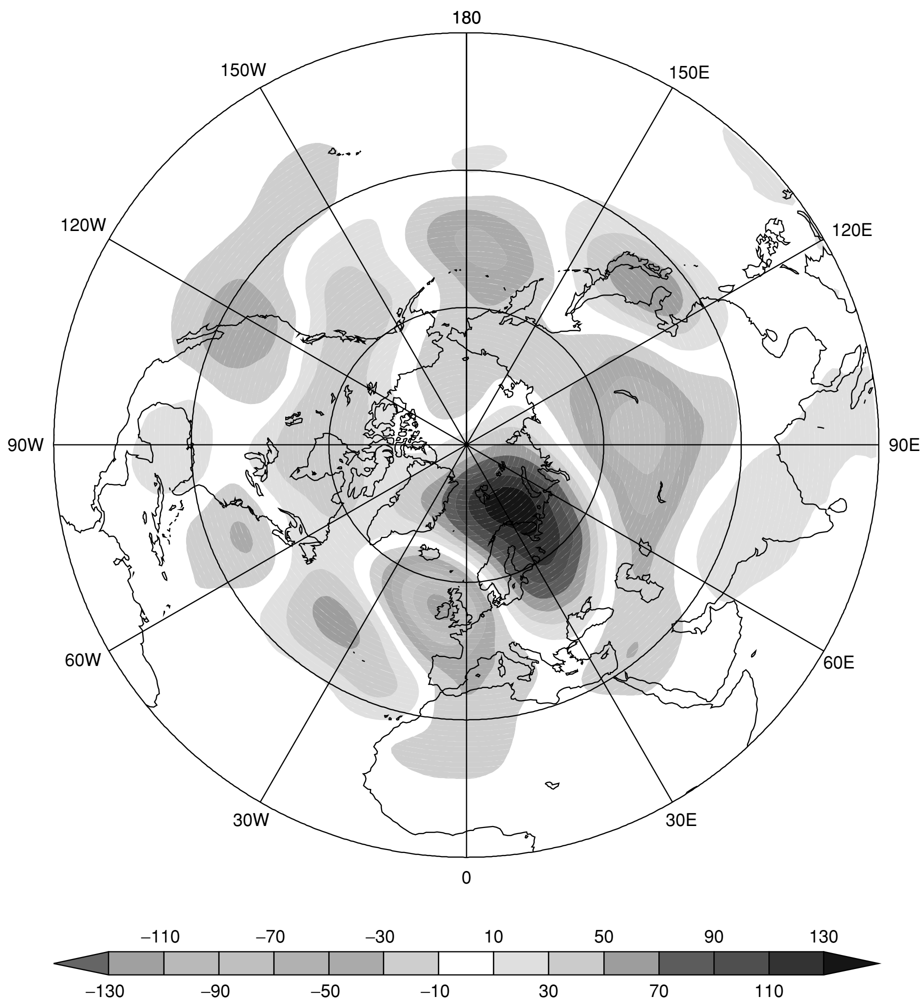

Figure 2.4 (See also colour plate section.) September–November 2000 300hPa geopotential height anomalies from climate. From Blackburn and Hoskins (personal communication).

On a β-plane Rossby wave activity spreads eastwards along regions of large meridional PV gradients as these provide the restoring mechanism for them. However, on the sphere, the path of long wavelength quasi-stationary Rossby wave activity is closer to great circles. Such Rossby waves generated in the tropics can arc polewards and eastwards into middle latitudes. Figure 2.4 (colour plate) shows the 300 hPa geopotential height anomalies for October 2000. As discussed by M. Blackburn and B. J. Hoskins (personal communication), the record rainfall over the UK is associated with the anomalous low over and to the west of the UK. It can be seen that this cyclone is preceded upstream, to the south-west, by an anticyclone. Streamfunction

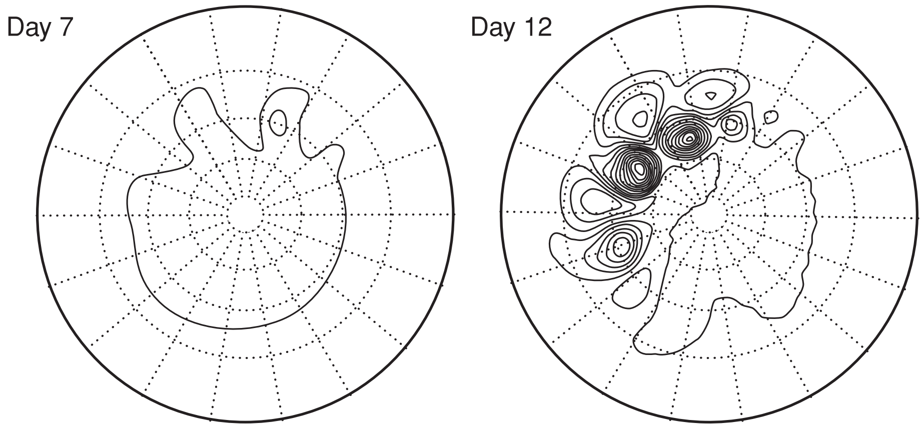

Figure 2.5 Surface pressure fields 7 days and 12 days after a baroclinically unstable zonal flow was disturbed near ‘3 o’clock’. The contours are drawn every 4 hPa. From Simmons and Hoskins (1979).

anomalies indicate a further cyclone south-west of this. The strong suggestion is that anomalous rainfall in the South American–Caribbean region triggered a stationary Rossby wave pattern that in turn led to the anomalous weather in the UK. If this hypothesis is correct, there is potential predictability of such anomalous midlatitude months. However, there must be the ability to represent the tropical anomalies in large-scale rainfall if it is to be realised.

2.3 Some phenomena

Middle latitude synoptic systems tend to have a characteristic structure including surface fronts and upper tropospheric troughs. They also have characteristic evolutions in time and therefore there is the implication of some predictability. These structural and evolutionary characteristics can be usefully interpreted in terms of theoretical descriptions in terms of normal mode baroclinic waves (Charney, 1947; Eady, 1949), non-linear life cycles (Simmons and Hoskins, 1978), the omega and vorticity equations (Hoskins *et al*., 1978), and coupled mid-troposphere and surface Rossby waves (Heifetz *et al*., 2004). A link with Rossby wave behaviour is shown by the ordered downstream development of baroclinic instability illustrated in Figure 2.5. In this numerical experiment from Simmons and Hoskins (1979) an unstable westerly flow had been perturbed at day zero at ‘3 o’clock’. Successive lows and highs have developed to the east and occluded, with the newest systems at day 12 near ‘6 o’clock’. Each system moves at some 10 m s^{-1}. However, the downstream development propagates at nearer 30 m s^{-1}, like a synoptic scale Rossby wave on the upper tropospheric jet. The orderly downstream propagation of baroclinic wave activity has been documented in a number of studies (e.g.

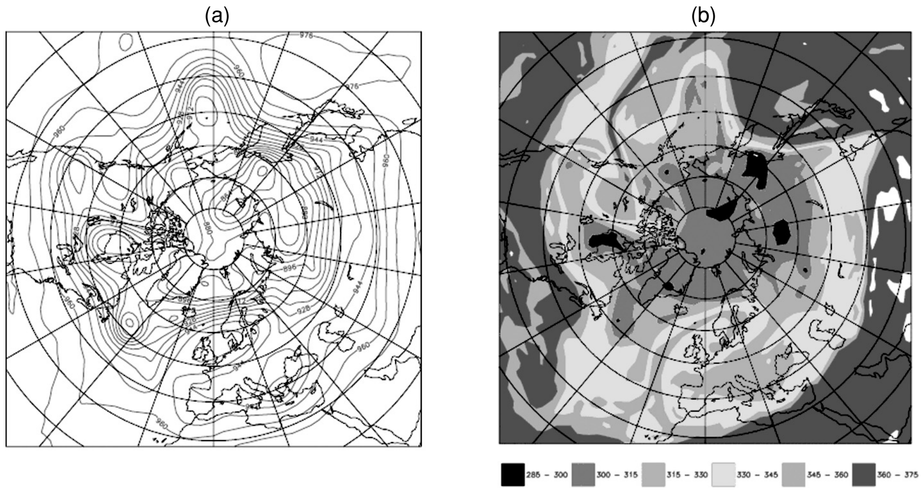

Figure 2.6 (See also colour plate section.) The block of 21 September 1998. Shown are the 250 hPa geopotential height field and the θ on PV2 field. From Pelly and Hoskins (2003a).

Chang, 1993). It gives the possibility that the development of new weather systems is predictable much beyond the synoptic timescale on which each individual system evolves.

The normal progression of middle latitude weather systems is sometimes interrupted by 'blocking highs'. An example of one of a block in the western European region is shown in Figure 2.6 (colour plate) in terms of its 300 hPa geopotential and θ on PV $= 2$ fields. The reversal of the zonal wind in the region of the block is associated with a reversal of the negative latitudinal gradient in θ. The formation of blocks, particularly in the European region, can often be viewed as a breaking of synoptic waves and, consistent with this, the timescale tends to be synoptic. However, once there is a low PV cut-off (here high θ cut-off) the decay is on the generally longer timescale of either diabatic processes or reabsorption into the subtropical region. Again there is associated enhanced predictability. These ideas are supported by Figure 2.7 from Pelly and Hoskins (2003a) which shows that on short timescales the decay of blocking-like features is on a timescale of about two days, but once a feature has lasted four days, and is probably associated with a PV cut-off, the decay time is about twice as long. It has indeed been found that the ECMWF Ensemble Prediction System has skill for the onset of blocking for about four days but for blocking events and the decay of them on timescales of about seven days (Pelly and Hoskins, 2003b).

A classic pattern of variability in the climate system is the North Atlantic Oscillation (NAO; Hurrell *et al.*, 2002) that describes the fluctuation of the surface westerly winds in that region. Alternative descriptions that have been used in recent years are the Arctic Oscillation (AO; Thompson and Wallace, 1998) that emphasises

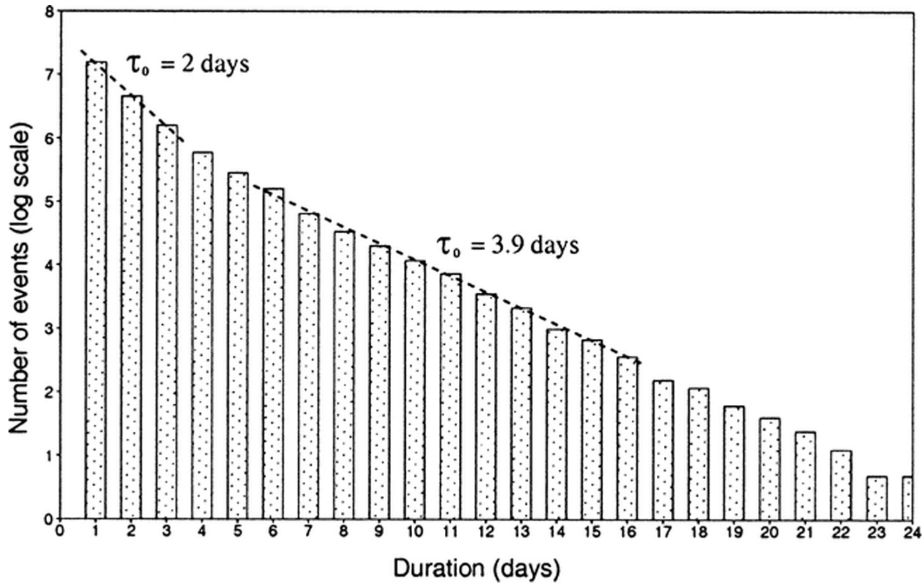

Figure 2.7 The number of blocking events in the period 1 August 2000 to 31 July 2001 lasting at least the number of days shown on the abscissa. The ordinate gives the logarithm of the number of sector blocking events and a straight line in the figure indicates a uniform decay rate of the events. For more details, see Pelly and Hoskins (2003a).

the fluctuations of the pressure in the polar cap and the Northern Annular Mode (NAM; Thompson and Wallace, 2000) that focuses on the polar vortex and the analogy with the southern hemisphere vortex. Figure 2.8 shows the autocorrelation timescales of the NAO and the northern hemisphere stratospheric vortex in winter (Ambaum and Hoskins, 2002). Synoptic timescale decay in the NAO changes to longer timescales beyond ten days. One hypothesis is that this reflects a link with the stratosphere and the longer timescales there, which are indicated in Figure 2.8.

The tropical Intra-Seasonal Oscillation (ISO), or Madden–Julian Oscillation (MJO), describes the large-scale organisation of tropical convection in the Indian Ocean that then migrates eastwards to the west Pacific, with its associated circulation changes. Consistent with the name, the timescale for the ISO is 30–60 days, although the word oscillation perhaps overemphasises its oscillatory nature. Once an ISO event has started, its typical evolution on the timescale of weeks is known. Since the convection associated with an ISO generates Rossby wave trains that lead to characteristic responses, particularly in the winter hemisphere (Matthews *et al.*, 2004), there is potential predictability on the timescale of weeks both in the tropics and higher latitudes. However, the current ability of models to simulate the ISO is generally poor and this potential predictability is yet to be realised.

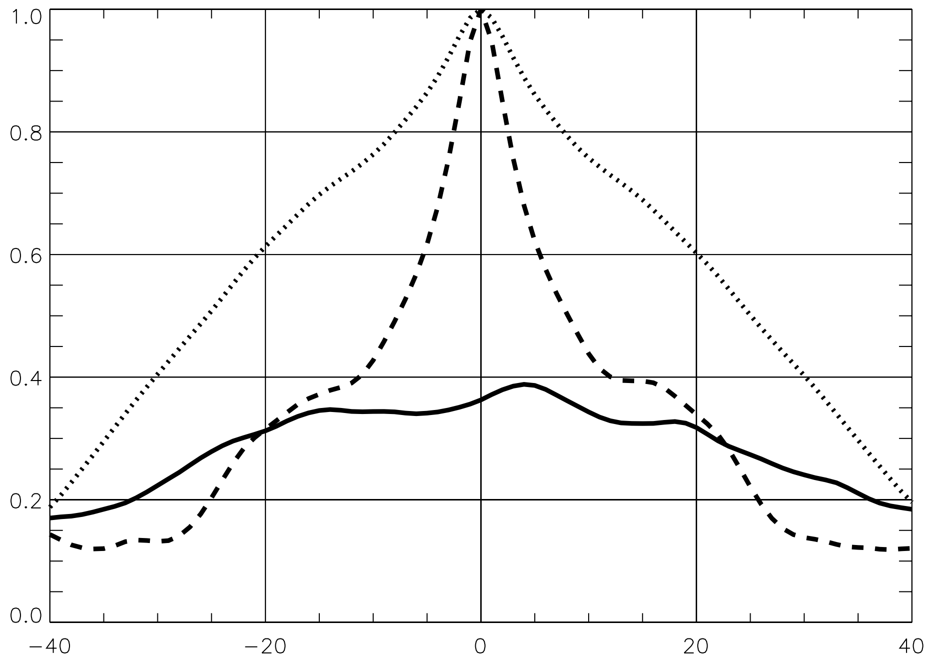

Figure 2.8 Lagged autocorrelations for the North Atlantic Oscillation (dashed) and the 500 K stratospheric vortex (dotted), and the lagged correlation between them (solid). The measures of the two patterns are the first principal components of daily mean sea-level pressure and 500 K PV, respectively. For more details, see Ambaum and Hoskins (2002).

2.4 Summer 2002

The northern summer of 2002 contained a number of climate features and anomalies that may have been linked. There were strong ISOs in the tropics. The surface westerly winds associated with these may have triggered the onset of an El Niño event that was observed to occur. Reduced Indian monsoon rainfall has been found to tend to occur in El Niño years and this certainly happened in 2002. The reduction in Indian monsoon rainfall was particularly dramatic in the middle of the season and was related to one very strong ISO event. The latent heat release associated with Indian monsoon rainfall leads to ascent in that region. It has been shown (Rodwell and Hoskins, 1996) that compensating descent occurs in the Mediterranean region and leads to the characteristic summer weather in the region. The reduction in the monsoon in 2002 is therefore consistent with the unusual occurrence in that year of weather systems moving into the Mediterranean region and then up into Europe, leading to flooding events there.

The possible linkages between all these events are explored by M. Blackburn *et al.* (personal communication). These linkages give the possibility of predictability.

However, they also indicate the range of processes and phenomena that may have to be modelled well in order to obtain this predictive power.

2.5 Concluding comments

A number of examples have been given in which phenomena and their structures give predictability. The dynamical perspective illustrated here provides a framework for consideration of the approach to prediction on various timescales and for the processes that need to be improved in models if potential predictability is to be found in practice.

References

Ambaum, M. H. P. and B. J. Hoskins (2002). The NAO troposphere-stratosphere connection. *J. Clim.*, **15**, 1969–78.

Chang, E. K. M. (1993). Downstream development of baroclinic waves as inferred from regression analysis. *J. Atmos. Sci*, **50**, 2038–53.

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.*, **4**, 135–62.

Eady, E. T. (1949). Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Heifetz, E., C. H. Bishop, B. J. Hoskins and J. Methven (2004). The counter-propagating Rossby-wave perspective on baroclinic instability. I: Mathematical basis. *Quart. J. Roy. Meteor. Soc.*, **130**, 211–31.

Hoskins, B. J. (1997). A potential vorticity view of synoptic development. *Meteorol. Appl.*, **4**, 325–34.

Hoskins, B. J., I. Draghici and H. C. Davies (1978). A new look at the ω-equation. *Q. J. Roy. Meteor. Soc.*, **104**, 31–8.

Hoskins, B. J., M. E. McIntyre and A. W. Robertson (1985). On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111**, 877–946.

Hurrell, J. W., Y. Kushnir, G. Ottersen and M. Visbeck (eds.) (2002). *The North Atlantic Oscillation: Climate Significance and Environmental Impact. Geophysical Monograph*, No. 134, American Geophysical Union.

Matthews, A. J., B. J. Hoskins and M. Masutani (2004). The global response to tropical heating in the Madden-Julian oscillation during the northern winter. *Quart. J. Roy. Meteor. Soc.*, **130**, 1991–2011.

Pelly, J. L. and B. J. Hoskins (2003a). A new perspective on blocking. *J. Atmos. Sci.*, **60**, 743–55.

(2003b). How well does the ECMWF Ensemble Prediction System predict blocking? *Quart. J. Roy. Meteor. Soc.*, **129**, 1683–702.

Rodwell, M. R. and B. J. Hoskins (1996). Monsoons and the dynamics of desserts. *Quart. J. Roy. Meteor. Soc.*, **122**, 1385–404.

Simmons, A. J. and B. J. Hoskins (1978). The life cycles of some non-linear baroclinic waves. *J. Atmos. Sci.*, **35**, 414–32.

(1979). The downstream and upstream development of unstable baroclinic waves. *J. Atmos. Sci.*, **36**, 1239–54.

Thompson, D. W. J. and J. M. Wallace (1998). The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–300.

(2000). Annular modes in the extratropical circulation. 1: Month-to-month variability. *J. Climate*, **13**, 1000–16.

3

Predictability – a problem partly solved

Edward N. Lorenz

Massachusetts Institute of Technology, Cambridge

Ed Lorenz, pioneer of chaos theory, presented this work at an earlier ECMWF workshop on predictability. The paper, which has never been published externally, presents what is widely known as the Lorenz 1996 model. Ed was unable to come to the 2002 meeting, but we decided it would be proper to acknowledge Ed's unrivalled contribution to the field of weather and climate predictability by publishing his 1996 paper in this volume.

The difference between the state that a system is assumed or predicted to possess, and the state that it actually possesses or will possess, constitutes the *error* in specifying or forecasting the state. We identify the rate at which an error will typically grow or decay, as the range of prediction increases, as the key factor in determining the extent to which a system is predictable. The long-term average factor by which an infinitesimal error will amplify or diminish, per unit time, is the leading Lyapunov number; its logarithm, denoted by λ_1, is the leading Lyapunov exponent. Instantaneous growth rates can differ appreciably from the average.

With the aid of some simple models, we describe situations where errors behave as would be expected from a knowledge of λ_1, and other situations, particularly in the earliest and latest stages of growth, where their behaviour is systematically different. Slow growth in the latest stages may be especially relevant to the long-range predictability of the atmosphere. We identify the predictability of long-term climate variations, other than those that are externally forced, as a problem not yet solved.

Predictability of Weather and Climate, ed. Tim Palmer and Renate Hagedorn. Published by Cambridge University Press.

3.1 Introduction

As I look back over the many meetings that I have attended, I recall a fair number of times when I have had the pleasure of being the opening speaker. It's not that this is necessarily a special honour, but it does allow me to relax, if not to disappear altogether, for the remainder of the meeting. On the present occasion, however, I find it is a true privilege to lead off. This is both because the subject of the seminar, predictability, is of special interest to me, and because much of the significant work in this field has taken place here at the European Centre.

Most of us who are here presumably have a special interest in the atmosphere, but the subject of predictability and the knowledge of it that we presently possess extend to much more general systems. By and large these systems fall into two categories, within which, to be sure, there are many subcategories. On the one hand there are real or realisable physical systems. On the other there are systems defined by mathematical formulas. The distinction between these categories is not trivial.

The former category includes the atmosphere, but also many much simpler systems, such as a pendulum swinging in a clock, or a flag flapping in a steady breeze. Instantaneous states of these systems cannot be observed with absolute precision, nor can the governing physical laws be expressed without some approximation. Exact descriptions of the dissipative processes are particularly elusive.

In the latter category, initial states may be prescribed exactly. Likewise, the defining formulas may be precisely written down, at least if the chosen finite-difference approximations to any differential equations, and the inevitable round-off procedures, are regarded as part of the system. In some instances the equations are of mathematical interest only, but in other cases they constitute models of real physical systems; that is, they may be fair, good, or even the best-known approximations to the equations that properly represent the appropriate physical laws. The relevance of mathematically defined systems cannot be too strongly emphasised; much of what we know, or believe that we know, about real systems has come from the study of models.

Systems whose future states evolve from their present states according to precise physical laws or mathematical equations are known as *dynamical systems*. These laws or equations encompass not only the internal dynamics of a system, but also any external factors that influence the system as it evolves. Often the concept of a dynamical system is extended to include cases where there may be some randomness or uncertainty in the evolution process, especially when it is believed that the general behaviour of the system would hardly be changed if the randomness could be removed; thus, in addition to mathematical models and abstractions, many real physical systems will qualify. Stochastic terms sometimes are added to otherwise deterministic mathematical equations to make them simulate real-system behaviour more closely.

In the ensuing discussion I shall frequently assume that our system is the atmosphere and its surroundings – the upper layers of the oceans and land masses – although I shall illustrate some of the points with rather crude models. By regularly calling our system the 'atmosphere' I do not mean to belittle the importance of the non-atmospheric portions. They are essential to the workings of the atmospheric portions, and, in fact, prediction of oceanic and land conditions can be of interest for its own sake, wholly apart from any coupling to the weather.

A procedure for predicting the evolution of a system may consist of an attempt to solve the equations known or believed to govern the system, starting from an observed state. Often, if the states are not completely observed, it may be possible to infer something about the unobserved portion of the present state from observations of past states; this is what is currently done, for example, in numerical weather prediction (see, for example, Toth and Kalnay, 1993). At the other extreme, a prediction procedure may be completely empirical. Nevertheless, whatever the advantages of various approaches may be, no procedure can do better than to duplicate what the system does. Any suitable method of prediction will therefore constitute, implicitly if not explicitly, an attempt at duplication – an attempt to reproduce the *result* of marching forward from the present state.

When we speak of 'predictability', we may have either of two concepts in mind. One of these is intrinsic predictability – the extent to which prediction is possible if an optimum procedure is used. The other is practical predictability – the extent to which we ourselves are able to predict by the best-known procedures, either currently or in the foreseeable future. If optimum prediction consists of duplication, it would appear that imperfect predictability must be due to one or both of two conditions – inability to observe the system exactly, and inability to formulate a perfect forward-extrapolation procedure. The latter condition is certainly met if the laws involve some randomness, or if future external influences cannot be completely anticipated.

When we cannot determine an initial state of a system precisely, there are two possible consequences. The system may be convergent; that is, two or more rather similar states, each evolving according to the same laws, may become progressively more similar. In this event, a precise knowledge of the true initial state is clearly not needed, and, in fact, the governing laws need not be known, since empirical methods will perform as well as any others. When we predict the oceanic tides, for example, which we can do rather well years in advance, we do not start from the observed present state of the ocean and extrapolate forward; we base our prediction on known periodicities, or on established relations between the tides and the computable motions of the sun, earth, and moon.

If, instead, the system is divergent, so that somewhat similar states become less and less similar, predictability will be limited. If we have no basis for saying which, if any, of two or more rather similar states is the true initial state, the governing laws cannot tell us which of the rather dissimilar states that would result from marching forward from these states will be the one that will actually develop. As will be noted in more detail in the concluding section, any shortcoming in the extrapolation procedure will have a similar effect. Systems of this sort are now known collectively as *chaos*. In the case of the atmosphere, it should be emphasised that it may be difficult to establish the absence of an intrinsic basis for discriminating among several estimates of an initial state, and the consequent intrinsic unpredictability; some estimates that now seem reasonable to us might, according to rules that we do not yet appreciate, actually be climatologically impossible and hence rejectable, while others might, according to similar rules, be incompatible with observations of earlier states.

3.2 First estimates of predictability

Two basic characteristics of individual chaotic dynamical systems are especially relevant to predictability. One quantity is the leading Lyapunov number, or its logarithm, the leading Lyapunov exponent. Let us assume that there exists a suitable measure for the difference between any two states of a system – possibly the distance between the points that represent the states, in a multidimensional phase space whose coordinates are the variables of the system. If two states are infinitesimally close, and if both proceed to evolve according to the governing laws, the long-term average factor by which the distance between them will increase, per unit time, is the first Lyapunov number. More generally, if an infinite collection of possible initial states fills the surface of an infinitesimal sphere in phase space, the states that evolve from them will lie on an infinitesimal ellipsoid, and the long-term average factors by which the axes lengthen or shorten, per unit time, arranged in decreasing order, are the Lyapunov numbers. The corresponding Lyapunov exponents are often denoted by $\lambda_1, \lambda_2, \ldots$; a positive value of λ_1 implies chaos (see, for example, Lorenz, 1993). Unit vectors in phase space pointing along the axes of the ellipsoid are the Lyapunov vectors; each vector generally varies with time.

Our interest in pairs of states arises from the case when one member of a pair is the true state of a system, while the other is the state that is believed to exist. Their difference is then the *error* in observing or estimating the state, and, if the assumed state is allowed to evolve according to an assumed law, while the true state follows the true law, their difference becomes the *error* in prediction. In the meteorological community it has become common practice to speak of the *doubling time* for small errors; this is inversely proportional to λ_1 in the case where the assumed and true laws are the same.

The other quantity of interest is the size of the attractor; specifically, the average distance ρ between two randomly chosen points of the attractor. The attractor is simply the set of points representing states that will occur, or be approximated arbitrarily closely, if the system is allowed to evolve from an arbitrary state, and transient effects associated with this state are allowed to die out. Estimation of these quantities is fairly straightforward for mathematically defined systems – ordinarily ρ^2 is simply twice the sum of the variances of the variables – but for real systems λ_1 may be difficult to deduce.

The third quantity that would seem to be needed for an estimate of the range of acceptable predictability is the typical magnitude of the error in estimating an initial state, ostensibly not a property of the system at all, but dependent upon our observing and inference techniques. For the atmosphere, we have a fair idea of how well we now observe a state, but little idea of what to expect in the years to come. Even though we may reject the notion of a future world where observing instruments are packed as closely as today's city dwellings, we do not really know what some undreamed-of remote-sensing technique may some day yield. However, assuming the size of an initial error, taking its subsequent growth rate to be given by λ_1, and recognising that the growth should cease when the predicted and actual states become as far apart as randomly chosen states – when the error reaches *saturation* – we can easily calculate the time needed for the prediction to become no better than guesswork.

How good are such naive estimates? We can demonstrate some simple systems where they describe the situation rather well, at least on the average. One system is one that I have been exploring in another context as a one-dimensional atmospheric model, even though its equations are not much like those of the atmosphere. It contains the K variables $X_1, \ldots, X_K$, and is governed by the K equations

$$dX_k/dt = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F, \tag{3.1}$$

where the constant F is independent of k. The definition of X_k is to be extended to all values of k by letting X_{k-K} and X_{k+K} equal X_k, and the variables may be thought of as values of some atmospheric quantity in K sectors of a latitude circle. The physics of the atmosphere is present only to the extent that there are external forcing and internal dissipation, simulated by the constant and linear terms, while the quadratic terms, simulating advection, together conserve the total energy $(X_1^2 + \cdots + X_K^2)/2$. We assume that $K > 3$; the equations are of little interest otherwise. The variables have been scaled to reduce the coefficients in the quadratic and linear terms to unity, and, for reasons that will presently appear, we assume that this scaling makes the time unit equal to 5 days.

For very small values of F, all solutions decay to the steady solution $X_1 = \ldots = X_K = F$, while, when F is somewhat larger, most solutions are periodic, but for still larger values of F (dependent on K) chaos ensues. For $K = 36$ and $F = 8.0$, for example, λ_1 corresponds to a doubling time of 2.1 days; if F is raised to 10.0, the time drops to 1.5 days.

Figures 3.1 and 3.2(a) have been constructed with $K = 36$, so that each sector covers 10 degrees of longitude, while $F = 8.0$. We first choose rather arbitrary values of the variables, and, using a fourth-order Runge–Kutta scheme with a time step Δt of 0.05 units, or 6 hours, we integrate forward for 14 400 steps, or 10 years. We then

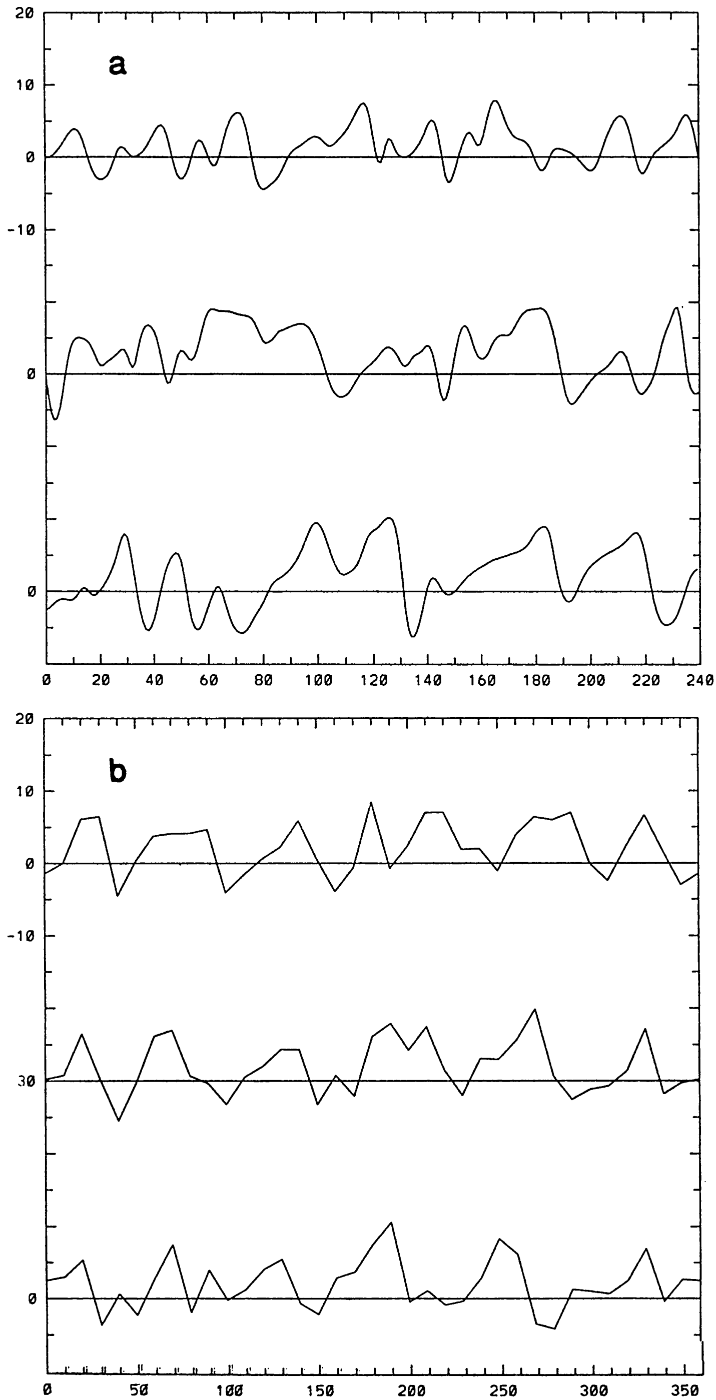

Figure 3.1 (a) Time variations of X_1 during a period of 180 days, shown as three consecutive 60-day segments, as determined by numerical integration of Eq. (3.1), with $K = 36$ and $F = 8.0$. Scale for time, in days, is at bottom. Scales for X_1 in separate segments are at left. (b) Longitudinal profiles of X_k at three times separated by 1-day intervals, determined as in (a). Scale for longitude, in degrees east, is at bottom. Scales for X_k in separate profiles are at left.

use the final values, which should be more or less free of transient effects, as new 'true' initial values, to be denoted by X_{k0}.

From Figure 3.1 we may gain some idea as to the resemblance or lack of resemblance between the *behaviour* of the model variables and some atmospheric variable such as temperature. Figure 3.1(a) shows the variations of X_1 during 720 time steps, or 180 days, beginning with the new initial conditions. The time series is displayed as three 60-day segments. There are some regularities – values lie mostly between -5 and $+10$ units, and about 12 maxima and minima occur every 60 days – but there is no sign of any true periodicity. Because of the symmetry of the model, all 36 variables should have statistically similar behaviour. Figure 3.1(b) shows the variations of X_k with k – a 'profile' of X_k about a 'latitude circle' – at the initial time, and one and two days later. The principal maxima and minima are generally identifiable from one day to the next, and they show some tendency to progress slowly westward, but their shapes are continually changing.

To produce the upper curve in Figure 3.2(a) we make an initial 'run' by choosing errors e_{k0} randomly from a distribution with mean 0 and standard deviation ε, here equal to 0.001, and letting $X'_{k0} = X_{k0} + e_{k0}$ be the 'observed' initial values of the K variables. We then use Eq. (3.1) to integrate forward from the true and also the observed initial state, for $N = 200$ steps, or 50 days, obtaining K sequences $X_{k0}, X_{k1}, \ldots, X_{kN}$ and K sequences $X'_{k0}, X'_{k1}, \ldots, X'_{kN}$, after which we let $e_{kn} = X'_{kn} - X_{kn}$ for all values of k and n.

We then proceed to make a total of $M = 250$ runs in the same manner, in each run letting the new values of X_{k0} be the old values of X_{kN} and choosing the values of e_{k0} randomly from the same distribution. Finally we let $e^2(\tau)$ be the average of the K values e^2_{kn}, where $\tau = n\Delta t$ is the prediction range, and let $\log E^2(\tau)$ be the average of the M values of $\log e^2(\tau)$, and plot $E(\tau)$ against the number of days (5τ), on a logarithmic scale. (The lower curve is the same except that the vertical scale is linear.)

For small n we see a nearly straight sloping line, representing uniform exponential growth, with a doubling time of 2.1 days, agreeing with λ_1, until saturation is approached. For large n we see a nearly straight horizontal line, representing saturation. It should not surprise us that the growth rate slackens before saturation is reached, rather than continuing unabated up to saturation and then ceasing abruptly.

The alternative procedure of simply letting $E^2(\tau)$ be the average value of $e^2(\tau)$, i.e. averaging the runs arithmetically instead of geometrically, would lead to a figure much like Figure 3.2(a), but with the sloping line in the upper curve indicating a doubling time of 1.7 days. Evidently the errors tend to grow more rapidly for a while in those runs where they have already acquired a large amplitude by virtue of their earlier more rapid growth, and it is these runs that make the major contribution to the arithmetic average. One could perhaps make equally good cases for studying geometric or arithmetic means, but only the former fits the definition of λ_1.

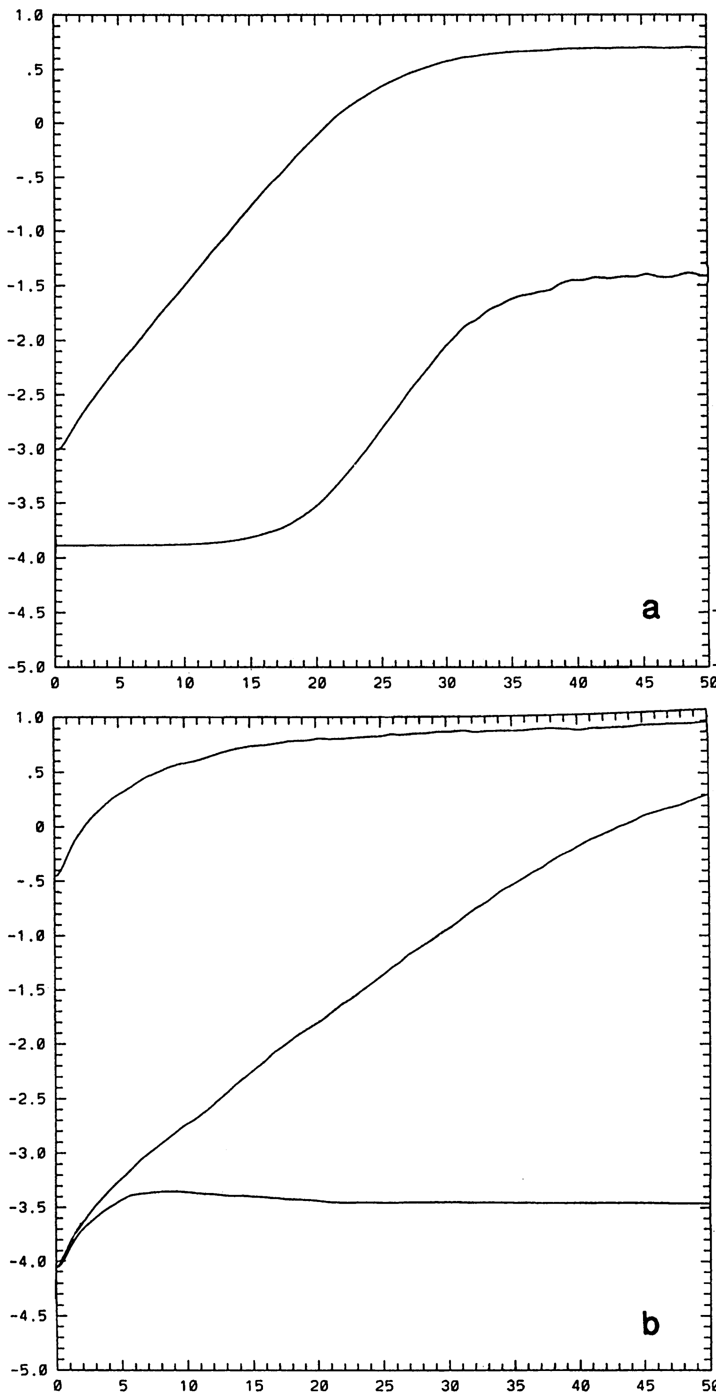

Figure 3.2 (a) Variations of average prediction error E (lower curve, scale at right) and $\log_{10} E$ (upper curve, scale at left) with prediction range τ (scale, in days, at bottom), for 50 days, as determined by 250 pairs of numerical integrations of Eq. (3.1), with $K = 36$ and $F = 8.0$ (as in Fig. 3.1). (b) The same as (a), but for variations of $\log_{10} E$ only, and as determined by 1000 pairs of integrations of Eq. (3.1), with $K = 4$, and with $F = 18.0$ (upper and middle curves, with different initial errors), and $F = 15.0$ (lower curve).

3.3 Atmospheric estimates

Some three decades ago a historic meeting, organised by the World Meteorological Organization, took place in Boulder, Colorado. The principal topic was long-range weather forecasting. At that time numerical modelling of the complete global circulation was just leaving its infancy; the three existing state-of-the-art models were those of Leith (1965), Mintz (1965), where A. Arakawa also played an essential role, and Smagorinsky (1965).

At such meetings the greatest accomplishments often occur between sessions. In this instance Jule Charney, who headed a committee to investigate the feasibility of a global observation and analysis experiment, persuaded the creators of the three models, all of whom were present, to use their models for predictability experiments, which would involve computations somewhat like those that produced Figure 3.2(a). On the basis of these experiments, Charney's committee subsequently concluded that a reasonable estimate for the atmosphere's doubling time was five days (Charney *et al.*, 1966). Taken at face value, this estimate offered considerable hope for useful two-week forecasts but very little for one-month forecasts.

The Mintz–Arakawa model that had yielded the five-day doubling time was a two-layer model. Mintz's graphs showed nearly uniform amplification before saturation was approached; presumably they revealed the model's leading Lyapunov exponent, although not, as we shall see, the leading exponent for the real atmosphere. As time passed by and more sophisticated models were developed, estimates of the doubling time appeared to drop. Smagorinsky's nine-level primitive-equation model, for example, reduced the time to three days (Smagorinsky, 1969).

Experiments more than a decade later with the then recently established operational model of ECMWF, based upon operational analyses and forecasts, suggested a doubling time between 2.1 and 2.4 days for errors in the 500-millibar height field (Lorenz, 1982). In the following years the model was continually modified, in an effort to improve its performance, and the newly accumulated data presently pushed the estimate below two days. There were small but significant variations of predictability with the season and the hemisphere, and quantities such as divergence appeared to be considerably less predictable than 500-m height.

One of the most recent studies (Simmons *et al.*, 1995), again performed with the ECMWF model, has reduced the estimate to 1.5 days. It is worth asking why the times should continually drop. Possibly the poorer physics of the earlier models overestimated the predictability, but it seems likely that a major factor has been spatial resolution. The old Mintz–Arakawa model used about 1000 numbers to represent the field of one variable at one level; the present ECMWF model uses about 45 000. Errors in features that formerly were not captured at all may well amplify more rapidly than those in the grossest features.

As with the Mintz–Arakawa model, the doubling times of the recent models appear consistent with the values of λ_1 for these models. Obviously not all of them can indicate the proper value of the exponent for the real atmosphere, and presumably none of them does.

Our reason for identifying the time unit in the model defined by Eq. (3.1) with five days of atmospheric time is now apparent. With $K = 36$ and $F = 8.0$ or 10.0, and indeed with any reasonably large value of K and these values of F, the doubling time for the model is made comparable to the times for the up-to-date global circulation models.

3.4 The early stages of error growth

Despite the agreement between the error growth in the simple model, and even in some global circulation models, with simple first estimates, reliance on the leading Lyapunov exponent, in most realistic situations, proves to be a considerable oversimplification. By and large this is so because λ_1 is defined as the long-term average growth rate of a very small error. Often we are not primarily concerned with averages, and, even when we are, we may be more interested in shorter-term behaviour. Also, in practical situations the initial error is often not small.

Sometimes, for example, we are interested in how well we can predict on specific occasions, or in specific types of situation, rather than in some general average skill. For any particular initial state, the initial growth rate of a superposed error will be highly dependent on the form of the error – on whether, for example, it assumes its greatest amplitude in synoptically active or inactive regions. In fact, there will be one error pattern – in phase space, it is an error vector – that will *initially* grow more rapidly than any other. The form and growth rate of this vector will of course depend upon the state on which the error is superposed.

Likewise, the *average* initial or early growth rate of *randomly* chosen errors superposed on a particular initial state will depend upon that state. Indeed, the identification of situations in which the atmosphere is especially predictable or unpredictable – the prediction of predictability – and even the identifiability of such situations – the predictability of predictability – have become recognised as suitable subjects for detailed study (see Kalnay and Dalcher, 1987; Palmer, 1988).

Assuming, however, that we are interested in averages over a wide variety of initial states, the value of λ_1 may still not tell us what we want to know, particularly in the earliest or latest stages of growth. In fact, in some systems the average *initial* growth rate of randomly chosen errors systematically exceeds the Lyapunov rate (see, for example, Farrell, 1989).

This situation is aptly illustrated by the middle curve in Figure 3.2(b), which has also been produced from Eq. (3.1), in the same manner as Figure 3.2(a), but with

K reduced to 4 and F increased to 18.0, and with $\varepsilon = 0.0001$. Also, because so few variables are averaged together, we have increased M to 1000. Between about 6 and 30 days the curve has a reasonably uniform slope, which agrees with λ_1, and indicates a doubling time of 3.3 days, but during the first 3.3 days the average error doubles twice. Systems exhibiting anomalously rapid initial error growth are in fact not uncommon. Certainly there are practical situations where we are mainly interested in what happens during the first few days, and here λ_1 is not always too relevant.

This phenomenon, incidentally, is in this case not related to the chaotic behaviour of the model. The lower curve in Figure 3.2(b) is like the middle one, except that F has been reduced to 15.0, producing a system that is not chaotic at all. Again the error doubles twice during the first six days, but then it levels off at a value far below saturation. If ε had been smaller, the entire curve would have been displaced downward by a constant amount.

When the initial error is not particularly small, as is often the case in operational weather forecasting, λ_1 may play a still smaller role. The situation is illustrated by the upper curve in Figure 3.2(b), which has been constructed exactly as the middle curve, except that $\varepsilon = 0.4$, or 5% of saturation, instead of 0.001. The rapid initial error growth is still present, but, when after four days it ceases, saturation is already being approached. Only a brief segment between 4 and 8 days is suggestive of 3.3-day doubling.

The relevance of the Lyapunov exponent is even less certain in systems, such as more realistic atmospheric models or the atmosphere itself, where different features possess different characteristic time scales. In fact, it is not at all obvious what the leading exponent for the atmosphere may be, or what the corresponding vector may look like. To gain some insight, imagine a relatively realistic model that resolves larger scales – planetary and synoptic scales – and smaller scales – mesoscale motions and convective clouds; forget about the fact that experiments with a global model with so many variables would be utterly impractical with today's computational facilities. Convective systems can easily double their intensity in less than an hour, and we might suppose that an initial error field consisting only of the omission of one incipient convective cloud in a convectively active region, or improperly including such a cloud, would amplify equally rapidly, and might well constitute the error pattern with the greatest *initial* growth rate.

Yet this growth rate need not be long-term, because the local instability responsible for the convective activity may soon subside, whereupon the error will cease to grow, while new instability may develop in some other location. A pattern with convective-scale errors distributed over many regions, then, would likely grow more steadily even if at first less rapidly, and might more closely approximate the leading Lyapunov vector.

Since this reasoning is highly speculative, I have attempted to place it on a slightly firmer basis by introducing another crude model which, however, varies with two distinct time scales. The model has been constructed by coupling two systems, each of which, aside from the coupling, obeys a suitably scaled variant of Eq. (3.1). There are K variables X_k plus JK variables $Y_{j,k}$, defined for $k = 1, \ldots, K$ and $j = 1, \ldots, J$, and the governing equations are

$$dX_k/dt = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k - (hc/b)\sum_{j=1}^{J} Y_{j,k}, \tag{3.2}$$

$$dY_{j,k}/dt = -cbY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + (hc/b)X_k. \tag{3.3}$$

The definitions of the variables are extended to all values of k and j by letting X_{k-K} and X_{k+K} equal X_k, as in the simpler model, and letting $Y_{j,k-K}$ and $Y_{j,k+K}$ equal $Y_{j,k}$, while $Y_{j-J,k} = Y_{j,k-1}$ and $Y_{j+J,k} = Y_{j,k+1}$. Thus, as before, the variables X_k can represent the values of some quantity in K sectors of a latitude circle, while the variables $Y_{j,k}$, arranged in the order $Y_{1,1}, Y_{2,1}, \ldots, Y_{J,1}, Y_{1,2}, Y_{2,2}, \ldots, Y_{J,2}, Y_{3,1}, \ldots$, can represent the values of some other quantity in JK sectors. A large value of J implies that many of the latter sectors are contained in one of the former, and we may think of the variables $Y_{j,k}$ as representing a convective-scale quantity, while, in view of the form of the coupling terms, the variables X_k should represent something that favours convective activity, possibly the degree of static instability.

In our computations we have let $K = 36$ and $J = 10$, so that there are ten small sectors, each one degree of longitude in length, in one large sector, while $c = 10.0$ and $b = 10.0$, implying that the convective scales tend to fluctuate ten times as rapidly as the larger scales, while their typical amplitude is $1/10$ as large. We have let h, the coupling coefficient, equal 1.0, and we have advanced the computations in time steps of 0.005 units, or 36 minutes. Our chosen value $F = 10.0$ is sufficient to make both scales vary chaotically; note that coupling replaces direct forcing as a driver for the convective scales.

Figure 3.3 reveals some of the typical behaviour of the model, by showing the distribution of X_k and $Y_{j,k}$ about a latitude circle, at times separated by 2 days. There are seven active areas (X_k large), generally 30 or 40 degrees wide, that fluctuate in width and intensity as they slowly propagate westward, while the convective activity, which is patently strongest in the active areas, tends to propagate eastward (note the signs in the subscripts in the non-linear terms in Eq. 3.3), but rapidly dies out as it leaves an active area.

Figure 3.4 presents separate error-growth curves for the large and small scales. For computational economy we have averaged 25 runs rather than 250. The small-scale errors begin to amplify immediately, doubling every 6 hours or so and approaching saturation by the third day. This growth rate is compatible with the computed value of λ_1 for the model. Meanwhile, the large-scale errors begin to grow at a similar rate once the small-scale errors exceed them by an order of magnitude, the growth evidently resulting from the coupling rather than the dynamics internal to the large scales. After the small-scale errors are no longer growing, the large-scale errors

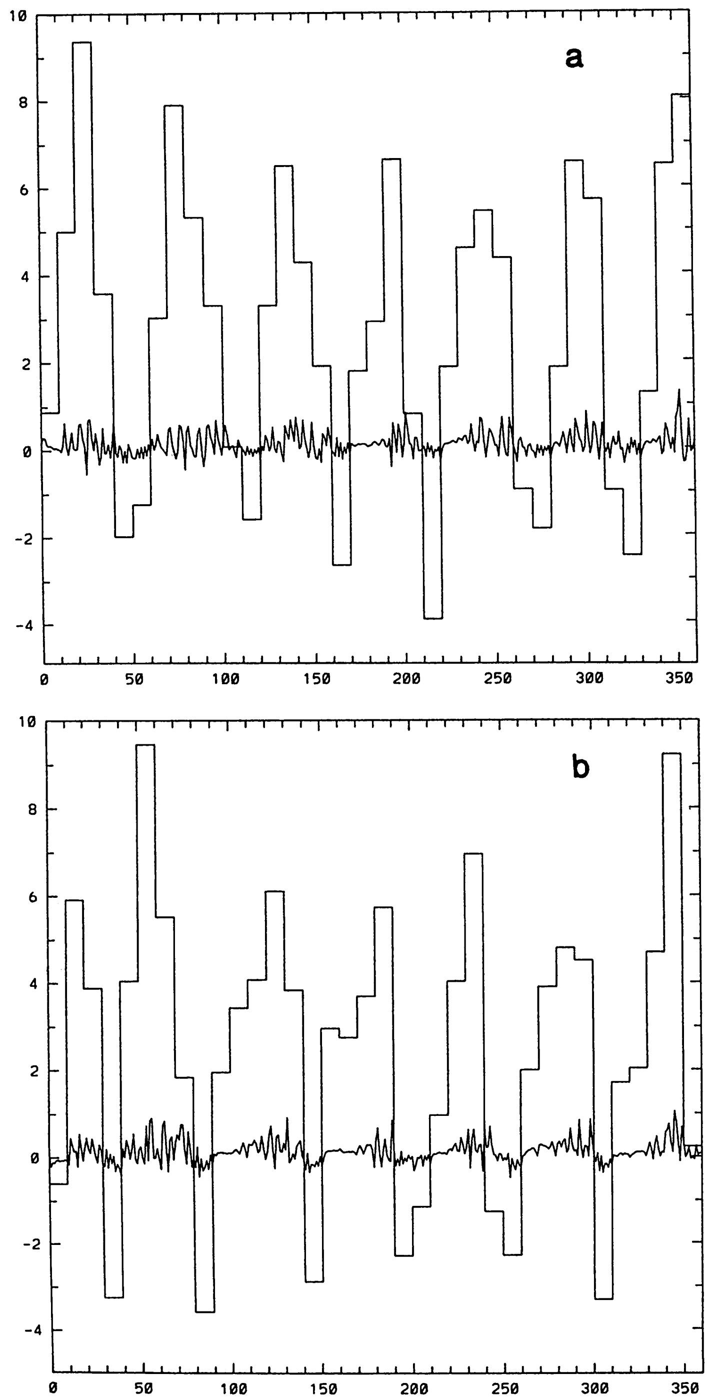

Figure 3.3 (a) Longitudinal profiles of X_k and $Y_{j,k}$ at one time, as determined by numerical integration of Eqs. (3.2) and (3.3), with $K = 36$, $J = 10$, $F = 10.0$, $c = 10.0$, $b = 10.0$, and $h = 1.0$. Scale for longitude, in degrees east, is at bottom. Common scale for X_k and $Y_{j,k}$ is at left. (b) The same as (a), but for a time two days later.

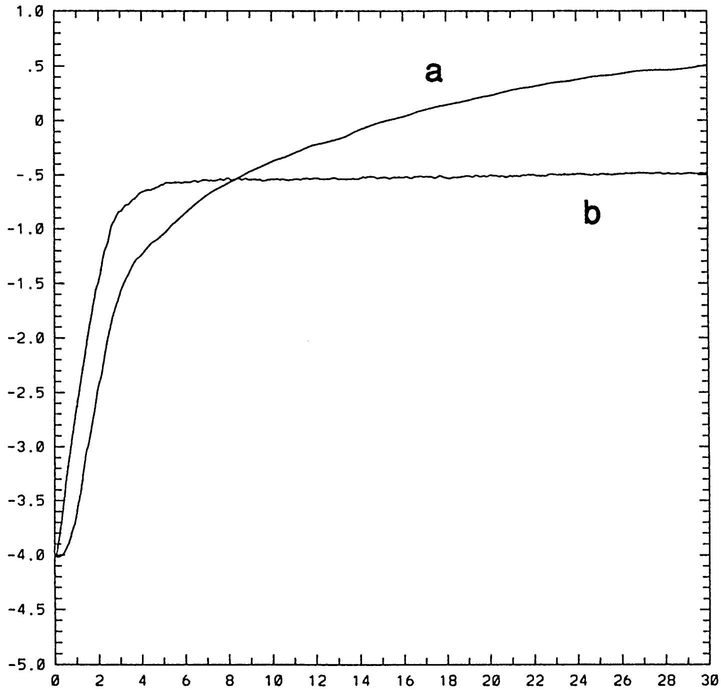

Figure 3.4 Variations of $\log_{10} E$ (scale at left) with prediction range τ (scale, in days, at bottom), shown separately for large scales (variables X_k, curve a) and small scales (variables $Y_{j,k}$, curve b), for 30 days, as determined by 25 pairs of integrations of Eqs. (3.2) and (3.3), with the parameter values of Figure 3.3.

continue to grow, at a slower quasi-exponential rate comparable to what appears in Figure 3.2(a), doubling in about 1.6 days. Finally they approach their own saturation level, an order of magnitude higher than that of the small-scale errors. Thus, after the first few days, the large-scale errors behave about as they would if the forcing were slightly weaker, and if the small scales were absent altogether.

In a more realistic model with many time scales or perhaps a continuum, we would expect to see the growth rate of the largest-scale errors subsiding continually, as, one after another, the smaller scales reached saturation. Thus we would not expect a large-scale-error curve constructed in the manner of Figure 3.4 to contain an approximate straight-line segment of any appreciable length.

We now see the probable atmospheric significance of the error doubling times of the various global circulation models. Each doubling time appears to represent the rate at which, in the *real* atmosphere, errors in predicting the features that are resolvable by the particular model will amplify, after the errors in unresolvable features have reached saturation. Of course, before accepting this interpretation, we must recognise the possibility that some of the small-scale features will not saturate rapidly; possibly they will act in the manner of coherent structures.

3.5 The late stages

As we have seen, prediction errors in chaotic systems tend to amplify less rapidly, on the average, as they become larger. Indeed, the slackening may become apparent long before the errors are close to saturation, and thus at a range when the predictions are still fairly good. For Eq. (3.1), and in fact for the average behaviour of some global atmospheric circulation models, we can construct a crude formula by assuming that the growth rate is proportional to the amount by which the error falls short of saturation. We obtain the equation

$$(1/E)dE/d\tau = \lambda_1(E^* - E)/E^*, \tag{3.4}$$

where E^* denotes the saturation value for E. Equation (3.4) possesses the solution

$$E = E^*(1 + \tanh(\lambda_1\tau))/2, \tag{3.5}$$

if the origin of τ is the range at which $E = E^*/2$. The well-known symmetry of the hyperbolic-tangent curve, when it is drawn with a linear vertical scale, then implies that the rate at which the error approaches saturation, as time advances, equals the rate at which it would approach zero, if time could be reversed. This relationship is evidently well approximated in the lower curve of Figure 3.2(a), and it has even been exploited to estimate growth rates for small errors, when the available data have covered only larger errors (see Lorenz, 1969b, 1982). It is uncertain whether the formula is more appropriate when E is the root-mean-square error or simply the mean-square error.

For many systems, however, Eq. (3.4) and hence Eq. (3.5) cannot be justified in the later stages. This may happen when, as in the case where the *early* growth fails to follow Eq. (3.4), the system possesses contrasting time scales. Here, however, the breakdown can occur because some significant feature varies more *slowly* than the features of principal interest – the ones that contribute most strongly to the chosen measure of total error.

Perhaps the feature most often cited as falling into this category is the sea surface temperature (SST), which, because of the ocean's high heat capacity, sometimes varies rather sluggishly. Along with the atmospheric features most strongly under its influence, the SST may therefore be expected to be somewhat predictable at a range when migratory synoptic systems are not. A slow final approach to saturation may thus be anticipated, particularly if the 'total error' includes errors in predicting the SST itself.

A perennial feature in which the SST plays a vital role is the El Niño–Southern Oscillation (ENSO) phenomenon. Phases of the ENSO cycle persist long enough for predictions of the associated conditions a few months ahead to be much better than guesswork, while some models of ENSO (e.g. Zebiak and Cane, 1987) suggest that the onsets of coming phases may also possess some predictability. Again, the phenomenon should lead to an ebbing of the late-stage growth rate.

Perhaps less important but almost certainly more predictable than the ENSO-related features are the winds in the equatorial middle-level stratosphere, dominated by the quasi-biennial oscillation (QBO). Even though one cannot be certain just when the easterlies will change to westerlies, or vice versa, nor how the easterlies or westerlies will vary from day to day within a phase, one can make a forecast with a fairly low expected mean-square error, for a particular day, a year or even several years in advance, simply by subjectively extrapolating the cycle, and predicting the average conditions for the anticipated phase. Any measure of the total error that gives appreciable weighting to these winds is forced to approach saturation very slowly in the latest stages.

Looking at still longer ranges, we come to the question, 'Is climate predictable?' Whether or not it is possible to predict climate changes, aside from those that result from periodic or otherwise predictable external activity, may depend on what is considered to be a climate change.

Consider again, for example, the ENSO phenomenon. To some climatologists, the climate changes when El Niño sets in. It changes again, possibly to what it had previously been, when El Niño subsides. We have already suggested that climatic changes, so defined, possess some predictability.

To others, the climate is not something that changes whenever El Niño arrives or leaves. Instead, it is something that often remains unchanged for decades or longer, and is characterised by the appearance and disappearance of El Niño at rather irregular intervals, but generally every two to seven years. A change of climate would be indicated if El Niño should start to appear almost every year, or only once in twenty years or not at all. Whether unforced changes of climate from one half-century or century to another, or one millennium to another, are at all predictable is much less certain.

Let us then consider the related question, 'Is climate a dynamical system?' That is, is there something that we can conscientiously call 'climate', determined by the state of the atmosphere and its surroundings, and undergoing significant changes over intervals of centuries but usually remaining almost unchanged through a single ENSO cycle or a shorter-period oscillation, whose future states are determined by its present and past according to some exact or approximate rule? To put the matter in perspective, let us first re-examine the justification for regarding the ever-changing synoptic pattern, and possibly the ENSO phenomenon, as dynamical systems.

Experience with numerical weather prediction has shown that we can forecast the behaviour of synoptic systems fairly well, far enough in advance for an individual storm to move away and be replaced by the next storm, without observing the superposed smaller-scale features at all, simply by including their influence in parametrised form. If instead of parametrising these features we omit them altogether, the models will still produce synoptic systems that behave rather reasonably, even though the actual forecasts will suffer from the omission. Evidently this is because the features that are small in scale are relatively small in amplitude, so that their influence acts much like small random forcing.

Moving to longer time scales, we find that some models yield rather good simulations of the behaviour of the ENSO phenomenon, even if not good forecasts of individual occurrences, without including the accompanying synoptic systems in any more than parametrised form. Here the synoptic systems do not qualify as being small in amplitude, but they appear to be rather weakly coupled to ENSO, so that again they may act like small random forcing.

Similarly, climatic fluctuations with periods of several decades or longer have more rapid oscillations superposed on them, ranging in timescale all the way from ENSO and the QBO to synoptic and small-scale features. Certainly these fluctuations are not small. Is their effect on the climate, if large, determined for the most part *by* the climate itself, so that climate can constitute a dynamical system? If this is not the case, are these features nevertheless coupled so weakly to the climate that they act like small random forcing, so that climate still constitutes a dynamical system? Or do they act more like strong random forcing, so that climate does not qualify as a dynamical system, and prospects for its prediction are not promising? At present the reply to these questions seems to be that we do not know.

3.6 Concluding remarks

In this overview I have identified the rate at which small errors will amplify as the key quantity in determining the predictability of a system. By an error we sometimes mean the difference between what is predicted and what actually occurs, but ordinarily we extend the concept to mean the difference, at any designated time, between two evolving states. We assume that there would be no prediction error if we could observe an initial state without error, and if we could formulate an extrapolation procedure without error, recognising that such formulation is not possible if the governing laws involve any randomness.

In my discussions and numerical illustrations I have found it convenient to consider the growth of errors that owe their existence to errors in the initial state, disregarding the additional influence of any inexactness in the extrapolation procedure. However, if the fault lies in the extrapolation and not in the initial state, the effect will be similar; after a reasonable time interval there will be noticeable errors in the *predicted* state, and these will proceed to grow about as they would have if they had been present initially. If the assumed and actual governing laws define systems with different leading Lyapunov exponents, the larger exponent will be the relevant one. Randomness in the governing laws will have the same effect as any other impediment to perfect