Editors
Ian T. Jolliffe | David B. Stephenson

Forecast Verification

A Practitioner's Guide in Atmospheric Science

SECOND EDITION

WILEY

Editors
Ian T. Jolliffe | David B. Stephenson

Forecast Verification

A Practitioner's Guide in Atmospheric Science

SECOND EDITION

WILEY

Contents

Cover

Title Page

Copyright

List of contributors

Preface

Preface to the first edition

1: Introduction

- 1.1 A brief history and current practice
- 1.2 Reasons for forecast verification and its benefits
- 1.3 Types of forecast and verification data
- 1.4 Scores, skill and value
- 1.5 Data quality and other practical considerations
- 1.6 Summary

2: Basic concepts

- 2.1 Introduction
- 2.2 Types of predictand
- 2.3 Exploratory methods

2.4 Numerical descriptive measures
2.5 Probability, random variables and expectations
2.6 Joint, marginal and conditional distributions
2.7 Accuracy, association and skill
2.8 Properties of verification measures
2.9 Verification as a regression problem
2.10 The Murphy–Winkler framework
2.11 Dimensionality of the verification problem

3: Deterministic forecasts of binary events

3.1 Introduction
3.2 Theoretical considerations
3.3 Signal detection theory and the ROC
3.4 Metaverification: criteria for assessing performance measures
3.5 Performance measures
Acknowledgements

4: Deterministic forecasts of multi-category events

4.1 Introduction
4.2 The contingency table: notation, definitions, and measures of accuracy
4.3 Skill scores

4.4 Sampling variability of the contingency table and skill scores

5: Deterministic forecasts of continuous variables

5.1 Introduction
5.2 Forecast examples
5.3 First-order moments
5.4 Second- and higher-order moments
5.5 Scores based on cumulative frequency
5.6 Summary and concluding remarks

6: Forecasts of spatial fields

6.1 Introduction
6.2 Matching methods
6.3 Traditional verification methods
6.4 Motivation for alternative approaches
6.5 Neighbourhood methods
6.6 Scale separation methods
6.7 Feature-based methods
6.8 Field deformation methods
6.9 Comparison of approaches
6.10 New approaches and applications: the future
6.11 Summary

7: Probability forecasts

7.1 Introduction
7.2 Probability theory

7.3 Probabilistic scoring rules
7.4 The relative operating characteristic (ROC)
7.5 Evaluation of probabilistic forecasting systems from data
7.6 Testing reliability
Acknowledgements

8: Ensemble forecasts

8.1 Introduction
8.2 Example data
8.3 Ensembles interpreted as discrete samples
8.4 Ensembles interpreted as probabilistic forecasts
8.5 Summary
Acknowledgement

9: Economic value and skill

9.1 Introduction
9.2 The cost/loss ratio decision model
9.3 The relationship between value and the ROC
9.4 Overall value and the Brier Skill Score
9.5 Skill, value and ensemble size
9.6 Applications: value and forecast users
9.7 Summary

10: Deterministic forecasts of extreme events and warnings

- 10.1 Introduction
- 10.2 Forecasts of extreme events
- 10.3 Warnings
- Acknowledgements

11: Seasonal and longer-range forecasts

- 11.1 Introduction
- 11.2 Forecast formats
- 11.3 Measuring attributes of forecast quality
- 11.4 Measuring the quality of individual forecasts
- 11.5 Decadal and longer-range forecast verification
- 11.6 Summary

12: Epilogue: new directions in forecast verification

- 12.1 Introduction
- 12.2 Review of key concepts
- 12.3 Forecast evaluation in other disciplines
- 12.4 Current research and future directions
- Acknowledgements

Appendix: Verification software

A.1 What is good software?
A.2 Types of verification users
A.3 Types of software and programming languages
A.4 Institutional supported software
A.5 Displays of verification information

Glossary

References

Plates

Index

Forecast Verification

A Practitioner's Guide in Atmospheric Science

SECOND EDITION

Edited by
Ian T. Jolliffe
David B. Stephenson
University of Exeter, UK

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2012
© 2012 by John Wiley & Sons, Ltd.

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information

in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Forecast verification : a practitioner's guide in atmospheric science / edited by Ian T. Jolliffe and David B. Stephenson. – 2nd ed.
p. cm.
Includes index.
ISBN 978-0-470-66071-3 (cloth)
1. Weather forecasting--Statistical methods--Evaluation. I. Jolliffe, I. T. II. Stephenson, David B.
QC996.5.F67 2011
551.63–dc23

2011035808

A catalogue record for this book is available from the British Library.

This book is published in the following electronic formats: ePDF 9781119960010; Wiley Online Library 9781119960003; ePub 9781119961079; Mobi 9781119961086

List of contributors

Dr Jochen Broecker
Max-Planck-Institute for the Physics of Complex Systems, Noethnitzer Str. 38, 01187 Dresden, Germany
broecker@pks.mpg.de
Dr Barbara G. Brown
Research Applications Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder CO 80307-3000, USA
bgb@ucar.edu
Michel Déqué
Météo-France CNRM,CNRS/GAME, 42 Avenue Coriolis, 31057 Toulouse Cedex 01, France
deque@meteo.fr
Dr Elizabeth E. Ebert
Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria 3001, Australia
e.ebert@bom.gov.au
Dr Christopher A.T. Ferro
Mathematics Research Institute, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK
c.a.t.ferro@exeter.ac.uk
Dr Eric Gilleland
Research Applications Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder CO 80307-3000, USA
ericg@ucar.edu
Professor Robin Hogan
Department of Meteorology, University of Reading, P.O. Box

243, Reading RG6 6BB, UK
r.j.hogan@reading.ac.uk
Professor Ian Jolliffe
30 Woodvale Road, Gurnard, Cowes, Isle of Wight, PO31 8EG, UK
ian@sandloch.fsnet.co.uk
Dr Robert E. Livezey
5112 Lawton Drive, Bethesda, MD 20816, USA
bobbilbo@msn.com
Dr Ian B. Mason
32 Hensman St., Latham, ACT, Australia, 2615
ibmason@grapevine.com.au
Dr Simon J. Mason
International Research Institute for Climate and Society (IRI), Columbia University, 61 Route 9W, P.O. Box 1000, Palisades, NY 10964-8000, USA
simon@iri.columbia.edu
Dr Matt Pocernich
Research Applications Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder CO 80307-3000, USA
e-mail: matt_pocernich@hotmail.com
Dr Jacqueline M. Potts
Biomathematics and Statistics Scotland, Craigiebuckler, Aberdeen AB15 8QH, UK
jackie@bioss.ac.uk
David S. Richardson
European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, RG2 9AX, UK
david.richardson@ecmwf.int
Professor David B. Stephenson
Mathematics Research Institute, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK
d.b.stephenson@exeter.ac.uk

Dr Andreas P. Weigel
Federal Office of Meteorology and Climatology MeteoSwiss, Kraehbuehlstr. 58, P.O. Box 514, CH-8044 Zurich, Switzerland
andreas.weigel@alumni.ethz.ch

Preface

In the eight years since the first edition was published, there has been considerable expansion of the literature on forecast verification, and the time is ripe for a new edition. This second edition has three more chapters than the first, as well as a new Appendix and substantially more references. Developments in forecast verification have not been confined to the atmospheric science literature but, as with the first edition, we concentrate mainly on this area.

As far as we are aware, there is still no other book that gives a comparable coverage of forecast verification, although at least two related books have appeared outside the atmospheric science area. Pepe (2003) is concerned with evaluation of medical diagnostic tests, which, although essentially concerned with 'forecast verification', has a very different emphasis, whilst Krzanowski and Hand (2009) is more narrowly focused on ROC curves.

We have retained many of the authors from the first edition, as well as bringing in a number of other experts, mainly for the new chapters. All are well-regarded researchers and practitioners in their fields. Shortly after the first edition was published, an extended and constructive review appeared (Glahn, 2004; Jolliffe and Stephenson, 2005). In this new edition we and our authors have attempted to address some of the issues raised by Glahn.

Compared with the first edition, the introductory and scene-setting Chapters 1 and 2 have only minor changes. Chapter 3 on 'Deterministic forecasts of binary events' has gained an additional author and has been rewritten. Much material from the first edition has been retained but has been restructured, and a non-trivial amount of new material, reflecting recent developments, has been added. Chapters 4

and 5 on, respectively, ‘Deterministic forecasts of multi-category events’ and ‘Deterministic forecasts of continuous variables’ have only minor improvements.

One of the biggest areas of development in forecast verification in recent years has been for spatial forecasts. This reflected by a much-expanded Chapter 6 on the topic, with three new authors, all of whom are leaders in the field.

In the first edition, probability forecasts and ensemble forecasts shared a chapter. This is another area of active development and, as suggested by Glahn (2004) and others, the two topics have been separated into Chapters 7 and 8 respectively, with two new authors. Chapter 9 on ‘Economic value and skill’ has only minor changes compared to the first edition.

Chapters 10 and 11 are both new, covering areas that have seen much recent research and are likely to continue to do so. Chapter 10 covers the related topics of verification of forecasts for rare and extreme events, and verification of weather warnings. By their nature the latter are often extreme, though many types of warnings are issued for events that are not especially rare. Impact rather than rarity is what warrants a warning. One context in which extremes are of particular interest is that of climate change. Because of the lack of verifying observations, the topic of verification of climate projections is still in its infancy, though likely to develop. There has been more activity on verification of seasonal and decadal forecasts, and these together with verification of climate projections, are the subject of Chapter 11.

The concluding Chapter 12 reviews some key concepts, summarizes some of the verification/evaluation activity in disciplines other than atmospheric sciences, and discusses some of the main developments since the first edition. As with the first edition, a Glossary is provided, and in addition there is an Appendix on available software. Although such

an Appendix inevitably becomes out of date more quickly than other parts of the text, it is arguably the most useful part of the book to practitioners for the first few years after publication. To supplement the Appendix, software and data sets used in the book will be provided via our book website: http://emps.exeter.ac.uk/fvb. We also intend to use this website to record errata and suggestions for future additions.

We hope you enjoy this second edition and find it useful. If you have any comments or suggestions for future editions, we would be happy to hear from you.

Ian T. Jolliffe
David B. Stephenson

Preface to the first edition

Forecasts are made in many disciplines, the best known of which are economic forecasts and weather forecasts. Other situations include medical diagnostic tests, prediction of the size of an oil field, and any sporting occasion where bets are placed on the outcome. It is very often useful to have some measure of the skill or value of a forecast or forecasting procedure. Definitions of 'skill' and 'value' will be deferred until later in the book, but in some circumstances financial considerations are important (economic forecasting, betting, oil field size), whilst in others a correct or incorrect forecast (medical diagnosis, extreme weather events) can mean the difference between life and death.

Often the 'skill' or 'value' of a forecast is judged in relative terms. Is forecast provider A doing better than B? Is a newly developed forecasting procedure an improvement on current practice? Sometimes, however, there is a desire to measure absolute, rather than relative, skill. Forecast verification, the subject of this book, is concerned with judging how good is a forecasting system or single forecast.

Although the phrase 'forecast **verification**' is generally used in atmospheric science, and hence adopted here, it is rarely used outside the discipline. For example, a survey of keywords from articles in the **International Journal of Forecasting** between 1996 and 2002 has no instances of 'verification'. This journal attracts authors from a variety of disciplines, though economic forecasting is prominent. The most frequent alternative terminology in the journal's keywords is 'forecast **evaluation**', although **validation** and **accuracy** also occur. Evaluation and validation also occur in other subject areas, but the latter is often used to denote a wider range of activities than simply judging skill or value – see, for example, Altman and Royston (2000).

Many disciplines make use of forecast verification, but it is probably fair to say that a large proportion of the ideas and methodology have been developed in the context of weather and climate forecasting, and this book is firmly rooted in that area. It will therefore be of greatest interest to forecasters, researchers and students in atmospheric science. It is written at a level that is accessible to students and to operational forecasters, but it also contains coverage of recent developments in the area. The authors of each chapter are experts in their fields and are well aware of the needs and constraints of operational forecasting, as well as being involved in research into new and improved methods of verification. The audience for the book is not restricted to atmospheric scientists – there is discussion in several chapters of similar ideas in other disciplines. For example ROC curves (Chapter 3) are widely used in medical applications, and the ideas of Chapter 8 are particularly relevant to finance and economics.

To our knowledge there is currently no other book that gives a comprehensive and up-to-date coverage of forecast verification. For many years, The WMO publication by Stanski **et al**. (1989) and its earlier versions was the standard reference for atmospheric scientists, though largely unknown in other disciplines. Its drawback is that it is somewhat limited in scope and is now rather out-of-date. Wilks (2006b [formerly 1995], Chapter 7) and von Storch and Zweirs (1999, Chapter 18) are more recent but, inevitably as each comprises only one chapter in a book, are far from comprehensive. The current book provides a broad coverage, although it does not attempt to be encyclopedic, leaving the reader to look in the references for more technical material.

Chapters 1 and 2 of the book are both introductory. Chapter 1 gives a brief review of the history and current practice in forecast verification, gives some definitions of

basic concepts such as skill and value, and discusses the benefits and practical considerations associated with forecast verification. Chapter 2 describes a number of informal descriptive ways, both graphical and numerical, of comparing forecasts and corresponding observed data. It then establishes some theoretical groundwork that is used in later chapters, by defining and discussing the joint probability distribution of the forecasts and observed data. Consideration of this joint distribution and its decomposition into conditional and marginal distributions leads to a number of fundamental properties of forecasts. These are defined, as are the ideas of accuracy, association and skill.

Both Chapters 1 and 2 discuss the different types of data that may be forecast, and each of the next five chapters then concentrates on just one type. The subject of Chapter 3 is binary data in which the variable to be forecast has only two values, for example {Rain, No Rain}, {Frost, No Frost}. Although this is apparently the simplest type of forecast, there have been many suggestions of how to assess them, in particular many different verification measures have been proposed. These are fully discussed, along with their properties. One particularly promising approach is based on signal detection theory and the ROC curve.

For binary data one of two categories is forecast. Chapter 4 deals with the case in which the data are again categorical, but where there are more than two categories. A number of skill scores for such data are described, their properties are discussed, and recommendations are made.

Chapter 5 is concerned with forecasts of continuous variables such as temperature. Mean squared error and correlation are the best-known verification measures for such variables, but other measures are also discussed including some based on comparing probability distributions.

Atmospheric data often consist of spatial fields of some meteorological variable observed across some geographical region. Chapter 6 deals with verification for such spatial data. Many of the verification measures described in Chapter 5 are also used in the spatial context, but the correlation due to spatial proximity causes complications. Some of these complications, together with some verification measures that have been developed with spatial correlation in mind, are discussed in Chapter 6.

Probability plays a key role in Chapter 7, which covers two topics. The first is forecasts that are actually probabilities. For example, instead of a deterministic forecast of ‘Rain’ or ‘No Rain’, the event ‘Rain’ may be forecast to occur with probability 0.2. One way in which such probabilities can be produced is to generate an ensemble of forecasts, rather than a single forecast. The continuing increase of computing power has made larger ensembles of forecasts feasible, and ensembles of weather and climate forecasts are now routinely produced. Both ensemble and probability forecasts have their own peculiarities that necessitate different, but linked, approaches to verification. Chapter 7 describes these approaches.

The discussion of verification for different types of data in Chapters 3–7 is largely in terms of mathematical and statistical properties, albeit properties that are defined with important practical considerations in mind. There is little mention of cost or value – this is the topic of Chapter 8. Much of the chapter is concerned with the simple cost-loss model, which is relevant for binary forecasts. However, these forecasts may be either deterministic as in Chapter 3, or probabilistic as in Chapter 7. Chapter 8 explains some of the interesting relationships between economic value and skill scores.

The final chapter (9) reviews some of the key concepts that arise elsewhere in the book. It also summarises the

aspects of forecast verification that have received most attention in other disciplines, including Statistics, Finance and Economics, Medicine, and areas of Environmental and Earth Science other than Meteorology and Climatology. Finally, the chapter discusses some of the most important topics in the field that are the subject of current research or that would benefit from future research.

This book has benefited from discussions and help from many people. In particular we would like to thank the following colleagues for their particularly helpful comments and contributions: Barbara Casati, Martin Goeber, Mike Harrison, Rick Katz, Simon Mason, Buruhani Nyenzi and Dan Wilks. Some of the earlier work on this book was carried out while one us (I.T.J.) was on research leave at the Bureau of Meteorology Research Centre (BMRC) in Melbourne. He is grateful to BMRC and its staff, especially Neville Nicholls, for the supportive environment and useful discussions; to the Leverhulme Trust for funding the visit under a Study Abroad Fellowship; and to the University of Aberdeen for granting the leave.

Looking to the future, we would be delighted to receive any feedback comments from you, the reader, concerning material in this book, in order that improvements can be made in future editions (see www.met.rdg.ac.uk/cag/forecasting).

1

Introduction

Ian T. Jolliffe and David B. Stephenson

Mathematics Research Institute, University of Exeter

Forecasts are almost always made and used in the belief that having a forecast available is preferable to remaining in complete ignorance about the future event of interest. It is important to test this belief *a posteriori* by assessing how skilful or valuable was the forecast. This is the topic of *forecast verification* covered in this book, although, as will be seen, words such as ‘skill’ and ‘value’ have fairly precise meanings and should not be used interchangeably. This introductory chapter begins, in Section 1.1, with a brief history of forecast verification, followed by an indication of current practice. It then discusses the reasons for, and benefits of, verification (Section 1.2). The third section provides a brief review of types of forecasts, and the related question of the target audience for a verification procedure. This leads on to the question of skill or value (Section 1.4), and the chapter concludes, in Section 1.5, with some discussion of practical issues such as data quality.

1.1 A brief history and current practice

Forecasts are made in a wide range of diverse disciplines. Weather and climate forecasting, economic and financial forecasting, sporting events and medical epidemics are

some of the most obvious examples. Although much of the book is relevant across disciplines, many of the techniques for verification have been developed in the context of weather, and latterly climate, forecasting. For this reason the current section is restricted to those areas.

1.1.1 History

The paper that is most commonly cited as the starting point for weather forecast verification is Finley (1884). Murphy (1996a) notes that although operational weather forecasting started in the USA and Western Europe in the 1850s, and that questions were soon asked about the quality of the forecasts, no formal attempts at verification seem to have been made before the 1880s. He also notes that a paper by Köppen (1884), in the same year as Finley's paper, addresses the same binary forecast set-up as Finley (see Table 1.1), though in a different context.

Table 1.1 Finley's tornado forecasts

	Observed		
Forecast	Tornado	No Tornado	Total
Tornado	28	72	100
No tornado	23	2680	2703
Total	51	2752	2803

Finley's paper deals with a fairly simple example, but it nevertheless has a number of subtleties and will be used in this and later chapters to illustrate a number of facets of forecast verification. The data set consists of forecasts of whether or not a tornado will occur. The forecasts were made from 10 March until the end of May 1884, twice daily, for 18 districts of the USA east of the Rockies. Table 1.1 summarizes the results in a table, known as a (2 × 2) contingency table (see Chapter 3). Table 1.1 shows that a total of 2803 forecasts were made, of which 100 forecast 'Tornado'. On 51 occasions tornados were observed, and on

28 of these ‘Tornado’ was also forecast. Finley's paper initiated a flurry of interest in verification, especially for binary (0–1) forecasts, and resulted in a number of published papers during the following 10 years. This work is reviewed by Murphy (1996a).

Forecast verification was not a very active branch of research in the first half of the twentieth century. A three-part review of verification for short-range weather forecasts by Muller (1944) identified only 55 articles ‘of sufficient importance to warrant summarization’, and only 66 were found in total. Twenty-seven of the 55 appeared before 1913. Due to the advent of numerical weather forecasting, a large expansion of weather forecast products occurred from the 1950s onwards, and this was accompanied by a corresponding research effort into how to evaluate the wider range of forecasts being made.

For the (2 × 2) table of Finley's results, there is a surprisingly large number of ways in which the numbers in the four cells of the table can be combined to give measures of the quality of the forecasts. What they all have in common is that they use the joint probability distribution of the forecast event and observed event. In a landmark paper, Murphy and Winkler (1987) established a general framework for forecast verification based on such joint distributions. Their framework goes well beyond the (2 × 2) table, and encompasses data with more than two categories, discrete and continuous data, and multivariate data. The forecasts can take any of these forms, but can also be in the form of probabilities.

The late Allan Murphy had a major impact on the theory and practice of forecast verification. As well as Murphy and Winkler (1987) and numerous technical contributions, two further general papers of his are worthy of mention here. Murphy (1991a) discusses the complexity and

dimensionality of forecast verification, and Murphy (1993) is an essay on what constitutes a ‘good’ forecast.

Weather and climate forecasting is necessarily an international activity. The World Meteorological Organization (WMO) published a 114-page technical report (Stanski *et al.*, 1989) that gave a comprehensive survey of forecast verification methods in use in the late 1980s. Other WMO documentation is noted in the next subsection.

1.1.2 Current practice

The WMO provides a Standard Verification System for Long-Range Forecasts. At the time of writing versions of this are available at a number of websites. The most up-to-date version is likely to be found through the link to the User's Guide on the website of the Lead Centre for the Long Range Forecast Verification System (http://www.bom.gov.au/wmo/lrfvs/users.shtml). The document is very thorough and careful in its definitions of long-range forecasts, verification areas (geographical) and verification data sets. It describes recommended verification strategies and verification scores, and is intended to facilitate the exchange of comparable verification scores between different centres. An earlier version is also available as attachments II-8 and II-9 in the WMO *Manual on the Global Data-Processing System* (http://www.wmo.int/pages/prog/www/DPS/Manual/WMO485.pdf). Attachment II-7 in the same document discusses methods used in standardized verification of NWP (Numerical Weather Prediction) products. Two further WMO documents can be found at http://www.wmo.int/pages/prog/amp/pwsp/pdf/TD-1023.pdf and http://www.wmo.int/pages/prog/amp/pwsp/pdf/TD-1103.pdf. These are respectively Guidelines (and Supplementary Guidelines) on Performance Assessment of Public Weather Services. The latter is discursive in nature,

whilst the guidelines in the former are more technical in nature.

European member states report annually on verification of ECMWF (European Centre for Medium Range Weather Forecasts) forecasts in their national weather services, and guidance on such verification is given in ECMWF Technical Memorandum 430 by Pertti Nurmi (http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/401-500/tm430.pdf).

At a national level, verification practices vary between different National Services, and most use a range of different verification strategies for different purposes. For example, verification scores used at the time of writing by the National Climate Centre at the Bureau of Meteorology in Australia range through many of the chapters that follow, for example proportion correct (Chapter 3), LEPS scores (Chapter 4), root mean square error (Chapter 5), anomaly correlation (Chapter 6), Brier skill score (Chapter 7) and so on (Robert Fawcett, personal communication).

There is a constant need to adapt practices, as forecasts, data and users all change. An increasing number of variables can be, and are, forecast, and the nature of forecasts is also changing. At one end of the range there is increasing complexity. Ensembles of forecasts, which were largely infeasible 30 years ago, are now commonplace (Chapter 8), and the verification of spatial forecasts has advanced significantly (Chapter 6). At the other extreme, a wider range of users requires targeted, but often simple (at least to express), forecasts. The nature of the data available with which to verify the forecasts is also evolving with increasing use of remote sensing by satellite and radar, for example.

An important part of any operational verification system is to have software to implement the system. As well as the widely available software described in Appendix, national

weather services often have their own systems. For example, the Finnish Meteorological Institute has a comprehensive operational verification package, which is regularly updated (Pertti Nurmi, personal communication).

A very useful resource is the webpage of the Joint Working Group on Forecast Verification Research (http://www.cawcr.gov.au/projects/verification/). It gives a good up-to-date overview of verification methods and issues associated with them, together with information on workshops and other events related to verification.

1.2 Reasons for forecast verification and its benefits

There are three main reasons for verification, whose description dates back to Brier and Allen (1951), and which can be described by the headings *administrative, scientific* and *economic*. Naturally no classification is perfect and there is overlap between the three categories. A common important theme for all three is that any verification scheme should be *informative*. It should be chosen to answer the questions of interest and not simply for reasons of convenience.

From an administrative point of view, there is a need to have some numerical measure of how well forecasts are performing. Otherwise, there is no objective way to judge how changes in training, equipment or forecasting models, for example, affect the quality of forecasts. For this purpose, a small number of overall measures of forecast performance are usually desired. As well as measuring improvements over time of the forecasts, the scores produced by the verification system can be used to justify funding for improved training and equipment and for research into

better forecasting models. More generally they can guide strategy for future investment of resources in forecasting.

Measures of forecast quality may even be used by administrators to reward forecasters financially. For example, the UK Meteorological Office currently operates a corporate bonus scheme, several elements of which are based on the quality of forecasts. The formula for calculating the bonus payable is complex, and involves meeting or exceeding targets for a wide variety of meteorological variables around the UK and globally. Variables contributing to the scheme range from mean sea level pressure, through precipitation, temperature and several others, to gale warnings.

The scientific viewpoint is concerned more with *understanding,* and hence improving the forecast system. A detailed assessment of the strengths and weaknesses of a set of forecasts usually requires more than one or two summary scores. A larger investment in more complex verification schemes will be rewarded with a greater appreciation of exactly where the deficiencies in the forecast lie, and with it the possibility of improved understanding of the physical processes that are being forecast. Sometimes there are unsuspected biases in either the forecasting models, or in the forecasters' interpretations, or both, which only become apparent when more sophisticated verification schemes are used. Identification of such biases can lead to research being targeted to improve knowledge of why they occur. This, in turn, can lead to improved scientific understanding of the underlying processes, to improved models, and eventually to improved forecasts.

The administrative use of forecast verification certainly involves financial considerations, but the third, 'economic', use is usually taken to mean something closer to the users of the forecasts. Whilst verification schemes in this case

should be kept as simple as possible in terms of communicating their results to users, complexity arises because different users have different interests. Hence there is the need for different verification schemes tailored to each user. For example, seasonal forecasts of summer rainfall may be of interest to both a farmer, and to an insurance company covering risks of event cancellations due to wet weather. However, different aspects of the forecast are relevant to each. The farmer will be interested in total rainfall, and its distribution across the season, whereas the insurance company's concern is mainly restricted to information on the likely number of wet weekends.

As another example, consider a daily forecast of temperature in winter. The actual temperature is relevant to an electricity company, as demand for electricity varies with temperature in a fairly smooth manner. In contrast, a local roads authority is concerned with the value of the temperature relative to some *threshold*, below which it should treat the roads to prevent ice formation. In both examples, a forecast that is seen as reasonably good by one user may be deemed ‘poor’ by the other. The economic view of forecast verification needs to take into account the economic factors underlying the users’ needs for forecasts when devising a verification scheme. This is sometimes known as ‘customer-based’ or ‘user-oriented’ verification, as it provides information in terms more likely to be understood by the ‘customer’ or ‘user’ than a purely ‘scientific’ approach. Forecast verification using economic value is discussed in detail in Chapter 9. Another aspect of forecasting for specific users is the extent to which users prefer a simple, less informative forecast to one that is more informative (e.g. a probability forecast) but less easy to interpret. Some users may be uncomfortable with probability forecasts, but there is evidence (Harold Brooks,