Contents

List of Contributors xi

Preface xiii

Preface to the First Edition xv

1 Introduction 1

Ian T. Jolliffe and David B. Stephenson

1.1 A brief history and current practice 1

1.1.1 History 1

1.1.2 Current practice 2

1.2 Reasons for forecast verification and its benefits 3

1.3 Types of forecast and verification data 4

1.4 Scores, skill and value 5

1.4.1 Skill scores 6

1.4.2 Artificial skill 6

1.4.3 Statistical significance 7

1.4.4 Value added 8

1.5 Data quality and other practical considerations 8

1.6 Summary 9

2 Basic concepts 11

Jacqueline M. Potts

2.1 Introduction 11

2.2 Types of predictand 11

2.3 Exploratory methods 12

2.4 Numerical descriptive measures 15

2.5 Probability, random variables and expectations 20

2.6 Joint, marginal and conditional distributions 20

2.7 Accuracy, association and skill 22

2.8 Properties of verification measures 22

2.9 Verification as a regression problem 23

2.10 The Murphy–Winkler framework 25

2.11 Dimensionality of the verification problem 28
CONTENTS

5.4.3 MSE of scaled forecasts 83
5.4.4 Correlation 84
5.4.5 An example: testing the ‘limit of predictability’ 86
5.4.6 Rank correlations 87
5.4.7 Comparison of moments of the marginal distributions 88
5.4.8 Graphical summaries 90
5.5 Scores based on cumulative frequency 91
  5.5.1 Linear Error in Probability Space (LEPS) 91
  5.5.2 Quantile-quantile plots 92
  5.5.3 Conditional quantile plots 92

5.6 Summary and concluding remarks 94

6 Forecasts of spatial fields 95

Barbara G. Brown, Eric Gilleland and
Elizabeth E. Ebert

6.1 Introduction 95
6.2 Matching methods 96
6.3 Traditional verification methods 97
  6.3.1 Standard continuous and categorical approaches 97
  6.3.2 S1 and anomaly correlation 98
  6.3.3 Distributional methods 99
6.4 Motivation for alternative approaches 100
6.5 Neighbourhood methods 103
  6.5.1 Comparing neighbourhoods of forecasts and observations 104
  6.5.2 Comparing spatial forecasts with point observations 104
6.6 Scale separation methods 105
6.7 Feature-based methods 108
  6.7.1 Feature-matching techniques 108
  6.7.2 Structure-Amplitude-Location (SAL) technique 110
6.8 Field deformation methods 111
  6.8.1 Location metrics 111
  6.8.2 Field deformation 112
6.9 Comparison of approaches 113
6.10 New approaches and applications: the future 114
6.11 Summary 116

7 Probability forecasts 119

Jochen Broecker

7.1 Introduction 119
7.2 Probability theory 120
  7.2.1 Basic concepts from probability theory 120
  7.2.2 Probability forecasts, reliability and sufficiency 121
7.3 Probabilistic scoring rules 122
  7.3.1 Definition and properties of scoring rules 122
  7.3.2 Commonly used scoring rules 124
  7.3.3 Decomposition of scoring rules 125
7.4 The relative operating characteristic (ROC) 126
CONTENTS

7.5 Evaluation of probabilistic forecasting systems from data 128
  7.5.1 Three examples 128
  7.5.2 The empirical ROC 130
  7.5.3 The empirical score as a measure of performance 130
  7.5.4 Decomposition of the empirical score 131
  7.5.5 Binning forecasts and the leave-one-out error 132

7.6 Testing reliability 134
  7.6.1 Reliability analysis for forecast A: the reliability diagram 134
  7.6.2 Reliability analysis for forecast B: the chi-squared test 136
  7.6.3 Reliability analysis for forecast C: the PIT 138

Acknowledgements 139

8 Ensemble forecasts 141
  Andreas P. Weigel
  8.1 Introduction 141
  8.2 Example data 142
  8.3 Ensembles interpreted as discrete samples 143
    8.3.1 Reliability of ensemble forecasts 144
    8.3.2 Multidimensional reliability 152
    8.3.3 Discrimination 157
  8.4 Ensembles interpreted as probabilistic forecasts 159
    8.4.1 Probabilistic interpretation of ensembles 159
    8.4.2 Probabilistic skill metrics applied to ensembles 160
    8.4.3 Effect of ensemble size on skill 163
  8.5 Summary 166

9 Economic value and skill 167
  David S. Richardson
  9.1 Introduction 167
  9.2 The cost/loss ratio decision model 168
    9.2.1 Value of a deterministic binary forecast system 169
    9.2.2 Probability forecasts 172
    9.2.3 Comparison of deterministic and probabilistic binary forecasts 174
  9.3 The relationship between value and the ROC 175
  9.4 Overall value and the Brier Skill Score 178
  9.5 Skill, value and ensemble size 180
  9.6 Applications: value and forecast users 182
  9.7 Summary 183

10 Deterministic forecasts of extreme events and warnings 185
  Christopher A.T. Ferro and David B. Stephenson
  10.1 Introduction 185
  10.2 Forecasts of extreme events 186
    10.2.1 Challenges 186
    10.2.2 Previous studies 187
    10.2.3 Verification measures for extreme events 189
CONTENTS

10.2.4 Modelling performance for extreme events 191
10.2.5 Extreme events: summary 194
10.3 Warnings 195
10.3.1 Background 195
10.3.2 Format of warnings and observations for verification 196
10.3.3 Verification of warnings 197
10.3.4 Warnings: summary 200
Acknowledgements 201

11 Seasonal and longer-range forecasts 203
Simon J. Mason

11.1 Introduction 203
11.2 Forecast formats 204
11.2.1 Deterministic and probabilistic formats 204
11.2.2 Defining the predictand 206
11.2.3 Inclusion of climatological forecasts 206
11.3 Measuring attributes of forecast quality 207
11.3.1 Skill 207
11.3.2 Other attributes 215
11.3.3 Statistical significance and uncertainty estimates 216
11.4 Measuring the quality of individual forecasts 217
11.5 Decadal and longer-range forecast verification 218
11.6 Summary 220

12 Epilogue: new directions in forecast verification 221
Ian T. Jolliffe and David B. Stephenson

12.1 Introduction 221
12.2 Review of key concepts 221
12.3 Forecast evaluation in other disciplines 223
12.3.1 Statistics 223
12.3.2 Finance and economics 225
12.3.3 Medical and clinical studies 226
12.4 Current research and future directions 228
Acknowledgements 230

Appendix: Verification Software 231
Matthew Pocernich

A.1 What is good software? 231
A.1.1 Correctness 232
A.1.2 Documentation 232
A.1.3 Open source/closed source/commercial 232
A.1.4 Large user base 232
A.2 Types of verification users 232
A.2.1 Students 233
A.2.2 Researchers 233
A.2.3 Operational forecasters 233
A.2.4 Institutional use 233