Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Abstracts of the Fifth International Verification Methods Workshop, 1 - 7 December 2011, Melbourne, Australia

CAWCR Technical Report No. 046

Keith A. Day (editor)

November 2011

www.cawcr.gov.au

Abstracts of the Fifth International Verification Methods Workshop, 1 - 7 December 2011 Melbourne, Australia

Keith A. Day (Editor)

Centre for Australian Weather and Climate Research,
GPO Box 1289, Melbourne, VIC 3001, Australia

CAWCR Technical Report No. 046

November 2011

Series ISSN: 1836-019X

National Library of Australia Cataloguing-in-Publication entry

Author: International Verification Methods Workshop (2011 : Melbourne, Victoria)

Title: Abstracts of the Fifth International Verification Methods Workshop, 1 - 7 December 2011, Melbourne, Australia / Editor Keith A. Day.

ISBN: 978-0-643-10730-4 (Electronic Resource – PDF)

Series: CAWCR technical report; No. 46

Notes: Includes index.

Subjects: Meteorology--Research--Congresses.

Enquiries should be addressed to:

Keith Day
Centre for Australian Weather and Climate Research:
A partnership between the Bureau of Meteorology and CSIRO
GPO Box 1289
Melbourne VIC 3001
Australia

k.day@bom.gov.au
Phone: 61 3 9669 8311
Fax: 61 3 9669 4660

Copyright and Disclaimer

Contents

Abstracts

Posters Day 1

Posters Day 2

FOREWORD

Forecast verification is an essential component of our science. It provides users with information on the accuracy and quality of forecasts, it enables us to judge whether a proposed forecasting system is better or worse than competing ones, and it provides error information that assists in improving the forecasts. The advent of new methods of forecasting, including ensemble prediction, very high resolution numerical weather prediction modelling, dynamical seasonal forecasts, and ocean forecasting, along with the continual improvement of observation networks, particularly from remote sensing, means that new approaches for forecast verification are not only possible but also necessary.

With these developments in mind, we welcome you to the Fifth International Verification Methods Workshop in Melbourne, Australia. This is the latest in a series of biannual workshops begun in 2002 to explore the many facets of how to evaluate forecasts. Previous workshops have been held in Boulder, Montreal, Reading, and Helsinki. The verification workshops are supported by the WMO World Weather Research Programme (WWRP), and coordinated by the WWRP/WGNE Joint Working Group on Forecast Verification Research (JWGFVR).

Since 2007 the workshop has also included a tutorial course that aims to teach and promote best verification practice, and build capacity in developing and developed nations. A benefit of holding this year's workshop in Australia is the opportunity to target participants from the Asia-Pacific region. We are pleased to have tutorial participants from 27 countries around the world, representing a great diversity of experience, forecasting issues, and verification needs.

This year's scientific workshop offers a stimulating and informative array of presentations on forecast verification relevant to a variety of topics including ensemble and probability forecasts, seasonal forecast and climate projections, aviation forecasts, user-focused verification, tropical cyclones and high impact weather forecasts and warnings, as well as methodologies and issues that apply to many areas. The workshop features five keynote talks by international experts, as well as a public lecture by renowned climate researcher, Professor Neville Nicholls of Monash University on, "Can weather and climate forecasts help avoid disasters?" Each topical session will conclude with an open discussion to capture the important issues and consider what research, data, and resources are necessary to advance that topic.

We hope that you will enjoy participating in the Fifth International Verification Methods Workshop, and that you will also take some time to enjoy the sights and sounds of the city of Melbourne.

Beth Ebert and Laurie Wilson
Co-chairs, WWRP/WGNE Joint Working Group on Forecast Verification Research

JWGFVR

VERIFICATION OF ENSEMBLE FORECASTS: A GUIDED TOUR THROUGH A ZOO OF SKILL METRICS

Andreas P. Weigel

Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland,
e-mail: andreas.weigel@meteoswiss.ch

A priori, ensembles are 'only' finite sets of deterministic forecast realizations that have been started from different initial conditions and/or are subject to different boundary conditions, and that are thought to represent samples from an underlying flow-dependent forecast probability distribution. In practice, ensembles are usually interpreted and applied as probabilistic forecasts, necessarily involving further statistical assumptions. These two levels of interpretation are reflected in the skill metrics commonly applied for ensembles, which can be categorized into two groups: on the one hand are tests that consider the individual ensemble members as discrete samples from a probability distribution, and on the other hand are truly probabilistic skill metrics that require that some form of probability distribution has been derived from the ensemble members prior to verification.

This talk will review types and techniques of ensemble verification for either of the two categories of ensemble interpretation. Both well-established and recently published skill metrics will be discussed, and recommendations and ideas for further research will be presented.

ENSEMBLE BASED PROBABILISTIC FORECAST VERIFICATION

Yuejian Zhu

Environmental Modeling Center, NCEP, NWS/NOAA, Washington DC,
e-mail: Yuejian.Zhu@noaa.gov

NCEP ensemble verification system has been developed to evaluate ensemble based probabilistic forecast since 90s. This system mainly focus on two attributes (reliability and resolution) for NCEP ensemble based probabilistic forecast in additional to the traditional measures, such as Pattern Anomaly Correlation (PAC) and Root Mean Square (RMS) error for ensemble mean, rank histogram, outlier and etc. The events definitions for probabilistic scores are based on 1) defined thresholds, 2) climatological percentiles and 3) defined by ensemble members. The probabilistic skill scores are based on NCEP/NCAR 40-year reanalysis climatology. Currently, this system generates Brier Scores (BS) and BS's decomposition (reliability and resolution), Ranked Probability Score (RPS), Continuous Ranked Probability Score (CRPS), Relative Operational Characteristics (ROC), Relative Economic Value (REV) and etc to apply to upper atmospheric variables, such as 500hPa height, 850hPa temperature, and near surface variables, such as 1000hPa height, 2-meter temperature, 10-meter winds, and precipitation. Since 2006, this system has been upgraded to apply for Northern American Ensemble Forecast System (NAEFS). Recently, this system will be applied to the future evaluation metrics for National Unified Operational Prediction Capability (NUOPC) implementation.

FROM VERIFICATION RESULTS TO PROBABILISTIC PRODUCTS: SPATIAL TECHNIQUES APPLIED TO ENSEMBLE FORECASTING

Zied Ben Bouallegue, Susanne Theis, Christoph Gebhardt

Deutscher Wetterdienst (DWD), Offenbach, Germany,
e-mail: zied.ben-bouallegue@dwd.de

Verification of precipitation forecasts derived from high resolution models is affected by the well known double penalty problem. In a point to point comparison, small displacements in space or time with regard to the observation penalize twice the forecast. Spatial techniques have been developed to alleviate this problem. In the framework of ensemble forecasting, information about the uncertainty of the prediction is delivered by means of probabilistic forecasts. In this context, the double penalty problem is no more an issue: the ensemble accounts for uncertainty in location. Nevertheless, it is still useful to apply spatial techniques in order to characterize better the potential of an ensemble forecast: its ability to describe the possible location of an event and its capacity to discriminate between event and not event at different spatial scales.

Two spatial techniques which take into account the spatial environment of each grid point forecast are investigated. The first one is the ‘neighbourhood’ method which enlarges the ensemble sample size to the neighbouring forecasts. ‘Fuzzy’ probabilistic forecasts are then produced for a range of neighbourhood sizes and verified focusing on the reliability gain. The second method is an upscaling procedure which modifies the reference area of the probabilities. Upscaled probabilistic forecasts are then generated for a range of spatial scales and verified focusing on the resolution gain.

Those two techniques are applied to the precipitation forecasts of the COSMO-DE-EPS, an ensemble prediction system based on the convection permitting model COSMO-DE which covers the area of Germany. Fuzzy and upscaled probabilistic forecasts interpretation and use are discussed.

VALIDATION, VERIFICATION AND VALUE: THE BRAVE NEW WORLD OF OPERATIONAL CLIMATE FORECASTING

Andrew B. Watkins

Climate Prediction Services, Bureau of Meteorology, Melbourne, Australia,
e-mail: a.watkins@bom.gov.au

In a day and age when anyone can place sea surface temperature and rainfall data into a spreadsheet, and without any knowledge of over-fitting or indeed atmospheric physics, derive a seasonal or longer range forecast, verification (assessing hindcast datasets) and validation (assessing independent outlook data) have become even more important when it comes to sorting the wheat from the chaff. But while the statistics of assessing seasonal forecasts provide a wealth of information for scientists and those technically inclined, the farmer or water manager or risk assessor want information that guides them as to the value of the seasonal forecasts. Recent work by the Bureau's Climate and Water Division has looked at some of the issues of delivering seasonal forecasts to the public, both how they interpret probabilities but also how they use verification information. Both results are somewhat startling; a significant proportion of users misinterpret odds, while many users not only struggled with verification information, many were not even aware the Bureau offered such a product. Improved communication, as well as using scores (such as reliability) that can be linked directly back to simple cost and loss, are ways to improve the understanding, and hopefully uptake, of data that gives an assessment of how valuable a seasonal forecast can be.

VERIFYING THE ACCURACY OF TWO DECADES OF SEASONAL CLIMATE OUTLOOKS

Harvey Stern and Jonathan Pollock

Bureau of Meteorology, Melbourne, Australia, e-mail: h.stern@bom.gov.au

Seasonal rainfall outlooks have been issued by the Bureau of Meteorology since the late 1980s, whilst seasonal minimum and maximum temperature outlooks have been issued from early 2000.

The purpose of this paper is to present preliminary results from an analysis of how accurate these outlooks have been. For almost the entire period, the outlooks have been represented by a map of Australia with the probability of the parameter predicted (total rainfall, mean minimum and mean maximum temperature) exceeding the median.

The verification methodology is now described:

For each State, season and prediction element, the forecast is regarded as equivalent to +1 should there be a region with >60% probability of exceeding the median and no region with <40% of exceeding the median.

For each State, season and prediction element, the forecast is regarded as equivalent to -1 should there be a region with <40% probability of exceeding the median and no region with >60% probability of exceeding the median.

For each State and three month season, all other forecasts are regarded as equivalent to 0. For each State, season and prediction element, the observed value is regarded as being equivalent to the anomaly.

The correlation coefficient between forecast and observed is calculated for each State, season and prediction element.

In summary, the correlation coefficients calculated on the data derived as described above are positive for most States, seasons and prediction elements, especially those for spring and summer rainfall outlooks. This is a very encouraging result.

WORKING GROUP ON SEASONAL TO INTERANNUAL PREDICTION (WGSIP)

Oscar Alves

CAWCR/Bureau of Meteorology, Melbourne, Australia, e-mail: o.alves@bom.gov.au

An overview of WGSIP projects will be given, with a particular focus on the WCRP Climate-system Historical Forecast Project (CHFP). This project is setting up a data server with a comprehensive collection of coupled model seasonal hind-casts in a consistent format. The dataset will be freely available to climate scientists and for use by various applications projects.

An overview of the techniques used for verifying seasonal forecasts will be given, particularly in relation to assessing the quality of forecasts from dynamical coupled models. Some of the issues faced will also be discussed.

CLIMATE PROJECTION EVALUATION: A PACIFIC CLIMATE CHANGE SCIENCE PROGRAM PERSPECTIVE

Damien Irving[1*], Sarah Perkins[1], Josephine Brown[2], Alex Sen Gupta[3], Aurel Moise[2], Bradley Murphy[2], Les Muir[1], Robert Colman[2], Scott Power[2], Francois Delage[2] and Jaclyn Brown[1]

[1]Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Australia, e-mail: damien.irving@csiro.au
[2]Centre for Australian Weather and Climate Research, Bureau of Meteorology, Australia
[3]Climate Change Research Centre, University of New South Wales, Sydney, Australia

Since the direct verification of any future climate projection is impossible, model agreement with observations of the present climate is used as a general approach to assigning model confidence. While this general approach is well established, there is much debate regarding the most appropriate methods of evaluating model simulations of the current climate. Recent work in this area has seen a move away from attempts to define a single metric of model skill, instead focusing on combining results from multiple metrics that assess many aspects of a climate model simulation. Consistent with this shift in thinking, the Pacific Climate Change Science Program (PCCSP) conducted a quantitative assessment of CMIP3 model performance in simulating the nine key aspects of a model simulation that are most relevant to the Pacific island region and the intended PCCSP projections. These included three climate variables ((1) surface air temperature, (2) precipitation and (3) surface wind), three climate features ((4) South Pacific Convergence Zone, (5) Intertropical Convergence Zone and the (6) West Pacific Monsoon), the (7) El Niño Southern Oscillation, (8) spurious model drift and (9) long term trends in sea surface temperature. There are few studies in the peer reviewed literature that have attempted to combine metrics pertaining to such a wide variety of climate processes and phenomena. In order to explore the practicalities of such an approach, this presentation will discuss in detail the PCCSP methodology and results.

In addition to issues surrounding the selection of appropriate model performance metrics, debate also surrounds the application of model evaluation results in determining climate projections, particularly with respect to the elimination and/or weighting of models on the basis of skill and the communication of climate projection uncertainty to the user community. This presentation will therefore also discuss the approaches taken by the PCCSP to incorporate model evaluation into the calculation and presentation of climate projections, including the development of a user-appropriate system for assigning a qualitative level of confidence to each projection.

EVALUATION OF HIGH RESOLUTION REGIONAL CLIMATE SIMULATIONS OVER THE ISLANDS IN THE SOUTH PACIFIC

Jack Katzfey and Mohar Chattopadhyay

CSIRO Marine and Atmospheric Research, Centre for Australian Weather and Climate Research (CAWCR), Aspendale, Australia, e-mail: jack.katzfey@csiro.au

The Pacific Climate Change Science Programme (PCCSP) was initiated by the Australian government in 2008 to provide better information on the likely impacts of climate change to the countries in the Pacific Ocean region. The climate change information can be provided by using the data from Global Climate Models (GCM)s. GCMs however simulate the atmosphere at a resolution of hundreds of kilometres which are relatively coarse when analysing the weather and climate of very small islands in the Pacific Ocean region. To overcome this limitation and to resolve the atmospheric processes at a finer resolution than the GCMs, a process to is applied to dynamically downscale the results from the GCMs. Dynamical downscaling uses large scale climate information from the GCMs as input to a regional climate model (RCM) which has more detailed information on topography, land use and coastal information.

One of the requirements for the PCCSP project was to dynamically downscale climates of seven Pacific Island nations at 8 km resolution using an RCM called CCAM and simulate the current and a future climate based on IPCC SRES A2 scenario using input from three GCMs. This study focuses on the evaluation and verification of the current climate as well as the future climate scenario simulated by CCAM.

The model results of temperature and precipitation are initially verified against gridded observations such as reanalysis and satellite data to understand whether the model can capture the intra-seasonal variations in these parameters. The results from the 8 km CCAM simulations have much more detail than that of the gridded data-sets. Hence the model grid-point data is compared with the station observations at various locations. Probability density function for various locations are also calculated and verified against the observations. All the above mentioned techniques are further applied to evaluate the model projections due a climate change scenario.

GLOBAL CLIMATE MODEL EVALUATION PLANS AT CAWCR

Aurel Moise[1], Damien Irving[2] and Penny Whetton[2]

[1]Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, Melbourne, Australia, e-mail: a.moise@bom.gov.au
[2]Centre for Australian Weather and Climate Research (CAWCR), CSIRO Marine and Atmospheric Research, Australia

A plan is underway to coordinate model evaluation across various CAWCR projects, in order to analyse the ability of the CMIP5 global climate models in simulating the climate of Australia and the western Pacific. A comparison of model and observed data using simple metrics (e.g. the root mean square error or pattern correlation of the climatological temperature or rainfall field) will be carried out in the first phase of the analysis, followed by a more detailed and complex second phase involving the analysis of various features of the modelled climate system. These features include: El Niño-Southern Oscillation (ENSO), South Pacific Convergence Zone (SPCZ), Inter-tropical Convergence Zone (ITCZ), Asian-Australian Monsoon, Sub-Tropical Ridge, Southern Annular Mode, Indian Ocean Dipole, west Pacific warm pool / cold tongue bias, Pacific Decadal Oscillation, climate extremes, synoptic-scale weather phenomena (e.g. fronts, blocking highs), and climate model drift.

A number of metrics and analysis methods have already been developed for this second phase of the analysis, including:

- Metrics for determining the mean latitude and orientation of the SPCZ and for assessing the relationship between ENSO and the mean SPCZ location (Brown et al., 2011).
- Metrics based on object-oriented pattern matching of rainfall (Moise & Delage, 2011). This allows for an assessment of the field structure with regard to errors in placement, rotation, volume and pattern.
- A metric for assessing the link between ENSO and interannual variability in the location of the ITCZ.
- Metrics for assessing ENSO, which capture (i) the strength and frequency of ENSO events, (ii) the mean climate and spatial pattern of ENSO, and (iii) the link between ENSO and climate variables such as precipitation (Irving et al., in press).
- Metrics for assessing the western Pacific monsoon, including the seasonality in both rainfall amount and wind reversal.
- Metrics for assessing climate model drift (Irving et al., in press).

This presentation/poster will give a more detailed overview of the planned CMIP5 model evaluation.

Brown JR., Power SB, Delage FP, Colman RA, Moise AF, Murphy BF (2011). Evaluation of the South Pacific Convergence Zone in IPCC AR4 climate model simulations of the 20th century. *Journal of Climate* 24: 1565-1582.

Moise AF, Delage FP (2011). New climate model metrics based on object-oriented pattern matching of rainfall. *Journal of Geophysical Research* 116: D12108. doi:10.1029/2010JD015318.

Irving DB, Perkins SE, Brown JR, Sen Gupta A, Moise AF, Murphy BF, Muir LC, Colman RA, Power SB, Delage FP, Brown JN (in press). Evaluating global climate models for climate change projections in the Pacific island region. *Climate Research*, doi:10.3354/cr01028.

RECENT DEVELOPMENTS IN AVIATION VERIFICATION AT THE UK MET OFFICE

Philip Gill

Met Office, Exeter,UK, e-mail: philip.gill@metoffice.gov.uk

Aviation remains one of the major customers for Met providers. Aviation verification plays an important role in ensuring safety of flight and in helping to reduce costs to the aviation industry.

The UK Met Office has recently undertaken a number of developments to improve the verification of aviation forecasts. In particular we look at three of these:

CAT verification

An objective scheme using high resolution automated aircraft observations to produce routine global verification for World Area Forecast Centre CAT forecasts. We discuss how this system is being used to help develop improved turbulence predictors.

Flight Time Error verification

An alternative measure of upper air wind forecast quality of direct relevance to airlines. Aircraft observations are compared with model wind forecasts. The flight time that would have been forecast for each flight can be calculated using model winds. The flight time error is the difference between the observed flight time and the forecast flight time.

TAF verification

A reliability table approach to give greater flexibility in assessing these probabilistic forecasts. A range of verification statistics can be produced from a simple web interface enabling forecasters to use the verification as a tool to improve the quality of their forecasts, to give external customers a measure of the value of the forecasts and to contribute to a corporate measure of organisational performance.

PMTIP: A USER BASED TREND FOR AVIATION PERFORMANCE METRICS

David Aihoshi

Environment Canada, Toronto, Canada, david.aihoshi@ec.gc.ca

This paper presents a summary on the Performance Measurement TAF Improvement Project (PMTIP). This is a new performance-based decision support system developed as a joint initiative between Environment Canada (EC), NAV CANADA, Air Canada, and Jazz Aviation. One of the goals for this project was to support air carrier decision making to decrease unnecessary fuel upload/carriage which would lower costs to aviation users and reduce related greenhouse gas emissions. Another goal was to create user-based performance metrics that are easy to use, view, and understand. PMTIP provides this for air carrier dispatchers by displaying new aviation performance metrics designed to relay essential information about the accuracy of EC's aerodrome forecasts which leads to increased user confidence in those forecasts. The metrics produced consist of critical ceiling and visibility categories, severe weather events, and wind as these were determined to be essential for flight planning by the air carrier representatives involved in the project. The development, methodology, and implementation of these new performance metrics available on PMTIP will be discussed.

IMPACT-BASED VERIFICATION: A BRIDGE BETWEEN WEATHER FORECAST AND OPERATIONAL DECISION

Jennifer Mahoney[1], Missy Petty[1,2], Geary Layne[1,3], Steve Lack[1,3], Matt Wandishin[1,3], Mike Kay[1,3], Andy Loughe[1,3], Brian Pettegrew[1,3], and Joan Hart[1,3]

[1]*NOAA/Earth System Research Laboratory/Global Systems Division, Boulder, CO, USA, e-mail:* Jennifer.Mahoney@NOAA.gov
[2]*Cooperative Institute for Research in the Atmosphere, Boulder, CO, USA*
[3]*Cooperative Institute for Research in Environmental Sciences, Boulder, CO USA*

Impact-based assessments and targeted information delivery are the bridge to improved operational decisions that are in response to high impact weather events. When forecast quality is measured in context to specific operational decisions, the forecast information is better utilized by the operational decision-maker. These concepts are illustrated here: three convective forecasts are used to strategically plan aviation flight routes for a particular day. One forecast is an automated high-resolution gridded forecast, the second is a human-generated polygon forecast, and the third is a probabilistic convective forecast. Each forecast is transformed to the operational grid of aviation sectors and decision time scales which are specific to the strategic planning process. The mincut-bottleneck methodology is applied to the three forecasts for assessing the potential impacts on the aviation sector space as predicted by the forecasts. Business rules determined by performance characteristics of the translated forecasts are used to formulate a decision matrix summarizing these potential impacts. This, along with continually-updating performance and impact information as events unfold are delivered to decision-makers and forecasters as input to the aviation-specific strategic plan.

As illustrated above, four operations are key to the process: transformation of the forecast methodologies, derivation of context-specific impact-based performance metrics, directed information delivery, and feedback to the forecast and decision processes. Our presentation will demonstrate this process by highlighting each operation with specific examples.

A TWO LAYER SCHEME TO VERIFY AERODROME FORECASTS (TAF)

Vivek C. Nagar

Australian Bureau of Meteorology, Melbourne, Australia, e-mail: v.nagar@bom.gov.au

The Bureau of Meteorology, in consultation with the aviation industry, has developed a new aviation forecast verification scheme (AVS2), the ultimate aim of which is an improvement in the forecast service. The new scheme tries to address most of the points raised in a review of the existing scheme (AVS1) carried out in March 2010.

The new scheme described in this paper has a two-layer structure. The first layer will be visible from within the Bureau and will store the detailed forecast and observation data. The second layer will be visible to the external clients and will display the forecast performance statistics through a user friendly web interface.

In the proposed scheme both short-term (previous shift/week) and long-term (previous month/year) verification results will be made available on demand. The new scheme will verify forecasts every hour. This means that for a 24 hour TAF, every forecast hour will be verified a minimum of four times because the TAF is issued every six hourly. In case of an amendment being issued for this TAF, every hour will be verified more than four times. In the first layer the scheme will create a contingency table for comparing forecasts and observations in different categories of weather elements interest e.g. ceiling, visibility etc. For the purpose of handling INTER and TEMPO forecasts a 60 minutes time window either sides of the forecast hour will be used. For the purpose of verifying TS, it is proposed that lightning strike data from GPATS and TSS928 sensors will be used. Both cloud-to-cloud and cloud-to-ground strikes within 5 nm of the airport will be used to verify TS.

The scheme will display statistics for different time slots of the forecast period, e.g. 0-6 hrs, 6-12 hrs, 18-24 hrs, 0-12 hrs etc. Similarly, verification statistics for selected UTC hours will be displayed, e.g. 20-22Z (for fog). It is proposed that the forecast verification results will be sent to the forecasters by email so that they can view their performance over last shifts and if necessary, consider a corrective strategy for the future.

FOG FORECASTING AT PERTH AIRPORT – APPLICATION OF A BAYESIAN DECISION NETWORK

Rodney Potts[1], *Tali Boneh*[1], *Yuelong Miao*[1], *Rabi Rivett*[2]

[1]*Centre for Australian Weather and Climate Research, Melbourne, Australia,*
e-mail: r.potts@bom.gov.au
[2]*Bureau of Meteorology, Perth, Australia*

The Bureau of Meteorology provides terminal aerodrome forecasts (TAF) for the aviation industry in Australia and when the probability of fog is 30% or greater this will be included in the TAF. In this case aircraft are required to carry additional fuel. If an unforecast fog occurs there are safety implications as there is a risk that an aircraft may not have sufficient fuel to reach a suitable alternate. Conversely, a false alarm has economic costs due to the carriage of unnecessary fuel and a reduction in load.

There are uncertainties in the forecasting of fog that can be difficult to quantify and the forecaster must consider these when assessing its probability. Because of these uncertainties and the safety implications there is a general tendency for forecasts of fog to be conservative to maintain a high probability of detection (POD). A result is that the frequency of hits (FOH) can be much lower than 30% and the false alarm ratio (FAR) can be high.

In recent years airlines have been developing more sophisticated strategies for managing operational risk and have indicated a requirement for more accurate forecasts of fog probability. With the aim of improving the forecast accuracy there has been work to develop a more structured fog forecasting process based on a Bayesian Decision Network (BDN). The focus of this work has been at Perth Airport on the SW coast of Australia.

The BDN was developed on past observational data and evaluated using cross validation techniques. The outcome of this was a list of contingency table results for a range of thresholds. In consultation with aviation industry representatives thresholds at three levels were selected to meet requirements for the POD, FAR and FOH. These were at a low level below which the likelihood of fog was very low. The highest threshold requires inclusion of fog in the TAF and this was selected to provide an acceptable POD with a FOH approaching 30%. At intermediate levels a Code Grey advice is issued indicating some risk of fog but with a nominal probability of 10%.

The presentation will give details on the BDN, including the performance evaluation, and present preliminary results from an operational trial currently underway at Perth Airport. Issues that have been identified with forecasters and aviation industry representatives and the implications of the shift in forecast practice will be presented.

UNCERTAINTY FORECASTS AND THE END-USER

Susan Joslyn

Department of Psychology, University of Washington, Seattle, USA,
e-mail: susanj@u.washington.edu

There is a growing trend toward adding uncertainty estimates to a wide array of weather variables in forecasts provided to general public end-users. Moreover, a growing body of psychological evidence suggests non-expert end-users can both understand such forecasts and benefit from them. This talk will provide an overview of a research program that investigates the benefits to decision-making as well as the principles for successfully communication of uncertainty forecasts. It is clear that uncertainty information is psychologically challenging. The reasons for that, as well as the common errors that arise, will be discussed. Several studies will be reported demonstrating the importance of communication format for easing processing load and reducing errors. Then, a series of studies will be described demonstrating that decisions based on simple uncertainty estimates, such as the probability of freezing, are better from an economic stand point than are decisions based on a deterministic forecast alone. Another group of studies will be reported demonstrating that users understand and benefit from more complex uncertainty forecasts called *predictive interval forecasts* and related verification graphics. The conclusion drawn from this research program is that uncertainty forecasts are clearly beneficial to everyday users, especially when the forecasts are carefully expressed.

USER-ORIENTED VERIFICATION: THE CONCEPT OF FORECASTER INTENT AND ITS APPLICATION TO VERIFICATION SYSTEMS

Timothy Bullock[1], *Ming Szeto*[2] *and Laurence Wilson*[3]

[1] *Science and Technology Branch, Environment Canada, Dartmouth, Canada,*
e-mail: Timothy.Bullock@ec.gc.ca
[2]*Meteorological Service of Canada, Dartmouth, Canada*
[3] *Science and Technology Branch, Environment Canada, Montréal, Canada*

Public forecast verification produced by the Meteorological Service of Canada is dismissed by many operational forecasters as irrelevant. Traditional verification methods for forecast regions are predicated on two assumptions:

1) at least one point in a region can be considered "representative";
2) instrumentation is installed at a "representative" point.

If either assumption is violated – as happens in many forecast regions in coastal areas and complex terrain - verification is compromised. This study proposes an approach to identify those regions with instrumented points that are sufficiently representative that traditional verification methods produce meaningful results.

To produce metrics meaningful to forecasters, we first determined what forecasters were trying to accomplish (i.e. their intent), then identified the forecast regions for which observations were suitable to measure the outcomes. Forecasters are well-positioned to advise on the utility of observing sites; their routine duties include continuous assessment of the observations. To guide the dialogue with forecasters, a framework was constructed, based on a simple model of their intent with three constituents: science, client needs, and organizational policies. To identify suitable observations, consideration was limited to: local climatology (science), distribution of population and roads (client need) and location of observing stations (policy). A preliminary analysis of these factors was undertaken for four public forecast weather parameters: temperature extrema, wind, cloud cover and precipitation, for each forecast region in Atlantic Canada. This analysis was presented to forecasters in a series of workshops at which consensus was achieved regarding the regions for which verification would be meaningful.

During the consultation process, forecasters identified:

- 17 regions for which temperature and wind observations are sufficiently representative,
- 15 regions for which either temperature or wind observations are sufficiently representative,
- two regions for which surface observations of cloud cover and precipitation are sufficiently representative.

Forecast verification is now provided for only these regions. Since forecaster acceptance of verification results has improved markedly, they are now used for Quality Management purposes. Examples of the Forecaster Intent analysis will be provided for selected regions in Atlantic Canada.

EXPECTED IMPACTS AND VALUE OF IMPROVEMENTS IN WEATHER FORECASTING ON THE TRANSPORTATION

Väinö Nurmi, Pertti Nurmi and Adriaan Perrels

Finnish Meteorological Institute, Helsinki, Finland, e-mail: pertti.nurmi@fmi.fi

This paper highlights the economic value of weather forecasts on transportation with particular emphasis on potential improvements in forecast quality in future climates. Atmospheric predictability has improved by one day per decade during the last 20 years based on ECMWF numerical forecasts. In Finland, the locally applied accuracy measures indicate a 5% improvement in forecast quality since late 1990s. There is no reason to question the continuation of such developments.

We first define the concept of economic value on micro level, where the value of information can be expressed in terms of increased expected utility and, thereafter, on macro level, where the value can be expressed as an increase in the social surplus. A simple cost-loss model shows that improved forecast accuracy leads to increased expected utility. This follows from decision-makers' ability to mitigate their losses more effectively. It is, however, not only the improved quality of available information that defines the value of weather information. The way information is communicated and how it is being used by decision-makers are equally important steps in the weather service value chain. This aspect will be emphasized with various examples. On a macro level, improved forecast quality leads to markets becoming less vulnerable to weather extremes. Capacity, representing the supply side of transportation, is reduced less than it would be without more accurate weather information. On the demand side travelers can make better pre-journey decisions and avoid accidents, delays and other costs caused by adverse weather conditions. The elasticity of short-run demand may increase because of increased public awareness

The transportation sector covers various sub-sectors and modes which are interlinked with each other on both supply and demand side of the markets. Potential mode substitution means that no sector or mode can be studied in isolation due to the network effect. There is no single methodology to estimate the value of weather information for all transport sectors because of high heterogeneity. Thus, we identify different valuation methods based on available literature and present applicable value estimates. Finally, some case studies are introduced to illustrate the use of these valuation methods.

VERIFICATION OF THE UK SHIPPING FORECAST USING A NEW OBJECTIVE AREA FORECAST VERIFICATION SYSTEM

Michael Sharpe

Met Office, Exeter, United Kingdom, email: Michael.sharpe@metoffice.gov.uk

Most Meteorological service providers issue text-based forecasts for geographical areas. Such area forecasts are particularly popular in marine products. The UK Met Office has recently developed a new automated Area Forecast Verification System with the initial purpose of objectively verifying its iconic Shipping Forecast - issued in the sea areas which surround the United Kingdom. Each shipping forecast is written by hand by a forecaster according to a predefined set of rules. An interpreter has been developed to turn the forecast text into a sequence of wind speed ranges. The Area Forecast Verification System compares each forecast wind speed range against the corresponding distribution formed by hourly nowcast analysis data in each sea area. The wind speed distribution is used to populate a contingency table of Beaufort Force categories from which a range of performance scores (including the Gerrity Score) are generated. Instant web based user feedback is provided to the forecaster in an attempt to improve the service - including a particularly useful graphical comparison between the forecast and the distribution of hourly nowcast data. By giving values to a set of parameters the customer is invited to tune the Gerrity score so that it reflects their particular priorities. The Area Forecast Verification System is also being used to verify other marine products and it is hoped that in future it will become the first port of call for the verification of most types of area forecast.

VERIFICATION PRESENTATION IN MANAGEMENT REPORTING

John Louey and *Milos Setek*

Bureau of Meteorology, Melbourne, Australia, e-mail: j.louey@bom.gov.au

The balance scorecard method of management reporting has evolved over the last fifteen years to be management best practice. The method leverages the benefits of strategic focus, complements the setting of proximal goals and harnesses organisational dynamics to improve performance.

Key Performance Indicators (KPIs), the products of this method, are now part of everyday business language. What do KPIs really mean? How might verification be presented to best unleash informational benefits?

The Bureau of Meteorology recently introduced a balanced scorecard design for executive reporting. Results for temperature forecast accuracy and rainfall accuracy were integrated as part of the scorecard design to provide accessible indication for the key goal of increasing forecast accuracy. The alchemy between verification science and management information design has resulted in a non-traditional presentation of traditional verification measures.

The session will provide participants with insights into management information needs and performance psychology by reviewing key ideas behind the balanced scorecard method, exploring the role of variability in management, and articulating how those factors have come to influence the Bureau's indicator design.

STATISTICAL ASSESSMENT OF TROPICAL CONVECTION-PERMITTING MODEL SIMULATIONS USING A CELL-TRACKING ALGORITHM

Simon Caine

School of Earth Sciences, University of Melbourne, Australia, e-mail: sclives@gmail.com

This study presents a method for comparing convection-permitting model simulations to radar observations using an object-based approach. The method uses the automated cell tracking algorithm, TITAN (Thunderstorm Identification Tracking Analysis and Nowcasting), to identify individual convective cells and determine their properties. Cell properties are identified in the same way for model and radar data, facilitating comparison of their statistical distributions. To demonstrate the capabilities of the method, it is applied to simulations of tropical convection during the Tropical Warm Pool - International Cloud Experiment using the Weather Research and Forecasting (WRF) model, and compared to data from a ground-based radar. Additional simulations with different microphysics and model resolution are also conducted. Among other things, the comparisons between the model and the radar elucidate model errors in the depth and size of convective cells. On average, simulated convective cells reached higher altitudes than the observations. Also, when using a low reflectivity (25 dBZ) threshold to define convective cells the model underestimates the size of the largest cells in the observed population. Some of these differences are alleviated with a change of microphysics scheme and higher model resolution, demonstrating the utility of this method for assessing model changes.

OBJECT-BASED SPATIAL VERIFICATION TECHNIQUES FOR DIAGNOSTIC ANALYSES OF PRECIPITATION PATTERNS DURING EXTREME RAINFALL EPISODES

Edward Tollerud[1], *Tara Jensen*[2], *Barbara Brown*[2], *Tressa Fowler*[2], *Wally Clark*[3], *Randy Bullock*[2], *John Halley Gotway*[2], *and Paul Oldenburg*[2]

[1]*Global Systems Division, NOAA/DOC and Developmental Testbed Center, Boulder, CO, USA,*
e-mail: edward.tollerud@noaa.gov
[2]*Research Applications Laboratory, NCAR and Developmental Testbed Center, Boulder, CO, USA*
[3]*Physical Sciences Division, NOAA/DOC, Boulder, CO, USA*

Extreme Rainfall events in the Sierra Nevada Mountains of California are often produced by narrow land-falling filaments (commonly termed atmospheric rivers or ARs) of very moist low- to mid-level air. The ability of numerical models to accurately predict the time and location of ARs and the subsequent inland rainfall they produce is a key factor in determining their value. Because standard pairs-based verification methods often do not give a helpful accounting of these forecasts, spatial techniques have been developed for the MET (Model Evaluation Tools) at NCAR that compare forecasts of precipitation and moisture flux objects to those observed. These tools are applied to heavy precipitation events captured by satellite and precipitation gage observations and predicted by a high-resolution WRF ensemble forecast system. Object attributes in forecast and observations including size, intensity, location, number, etc., are used to quantitatively compare model predictions from different research and operational models, and as diagnostic guidance to provide clues to differences in physical characteristics of the WRF ensemble members that pertain to the accuracy of their precipitation forecasts. As counter-point to these studies based on QPF (Quantitative Precipitation Forecasts), another use of objects based on reflectivity probability fields for convective rainfall in the central United States is presented. Finally, a method is suggested to use verification objects defined on a Hovmueller (time and space) diagram to assess the ability of numerical forecasts to simultaneously capture the timing and coastline location of the land-fall of ARs.

MORPHING FOR GRID EDITING AND VERIFICATION IN AN OPERATIONAL WEATHER FORECAST SYSTEM

G. T. Weymouth and N.Faggian

Bureau of Meteorology, Melbourne, Australia, e-mail: G.Weymouth@bom.gov.au

Aims

The Australian Bureau of Meteorology has expanded its services through introduction of a gridded Next Generation Forecast and Warning System (NexGenFWS). NexGenFWS automates many aspects of forecast production while supporting forecaster oversight and editing.

Morphing (or 'warping') is being developed within the NexGenFWS for aims including

1. manually-driven weather feature position adjustment
2. improved manually-driven or automated time interpolation that accounts for the motion of weather systems
3. high-resolution weather field verification to separately estimate displacement and magnitude errors.

We aim to support inter-field consistency of morphs. Central to the work is the verification of the approaches taken, including a novel approach to measure the utility of the manually-driven morphs. An open-source morphing project has been initiated to pool development resources.

Method – verification of manually-driven morphs

Hourly NWP guidance and analysis fields can either be morphed or considered the 'truth'. One (simplified) verification method studied is to treat spatial displacement within one grid (for example correction of the position of a cold front) as being approximately equivalent to feature motion over time:

1) estimate feature location in a wind field at time points *t* and *t*+1
2) morph the time *t* wind feature (plus other fields concurrently) to the position estimated at time *t*+1. Use appropriate methods to avoid morphing topographically-related features. For example, mountain-top temperatures should not morph unchanged into valleys.

3) compare against *t*+1 grid for appropriate fields:
   a) morphed *t* grid
   b) unmodified *t* grid
   c) linear time interpolation between *t*+0 and *t*+2 (etc)

Traditional comparison metrics such as RMS differences are used along with morphing distance-related measures.

Results and conclusion

Our results show that

- morphing interpolation and feature position adjustment are valid approaches
- these outperform linear time interpolation

- reasonable inter-field consistency and topographic effects are maintained when multiple fields are morphed at the same time

For example, using the method outlined for morphing position adjustment of surface temperature in 4 cold front cases, we obtain RMSE for the morphed = 0.79 * RMSE for the unmorphed grid at time *t*, compared to the *t*+1 grid 'truth'. Improvements are expected.

USING THE SPECTRAL SCALING EXPONENT FOR VALIDATION OF QUANTITATIVE PRECIPITATION FORECASTS

T.-Y. Koh[1,2,3], *B. C. Bhatt*[3], *K. K. W. Cheung*[4], *C. K. Teo*[3], *Y. H. Lee*[5], *Matthias Roth*[6], *Purnawirman*[3]

[1]*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, e-mail:* KohTY@ntu.edu.sg
[2]*Earth Observatory of Singapore, Nanyang Technological University, Singapore*
[3]*Temasek Laboratories, Nanyang Technological University, Singapore*
[4]*Department of Environment and Geography, Macquarie University, Australia, e-mail:* kevin.cheung@mq.edu.au
[5]*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*
[6]*Department of Geography, National University of Singapore, Singapore*

This study evaluates the spectral scaling of a heavy rainfall event and assesses the performance of Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) model in terms of the multiscale variability of rainfall in the temporal spectral domain. The event occurred over southern Malay Peninsula on 18 December 2006 and was simulated at high resolutions. 10, 5 and 1-minute aggregate rainfall data from rain gauge stations in Singapore and simulated rainfall sampled at different evaluation points on 0.9, 0.3 and 0.1 km grids were utilized. The simulated and observed rain rates were compared via Fourier and wavelet analyses. A scaling regime was noted in the observed rainfall spectra in the timescales between 60 min and 2 min. The scaling exponent obtained from the observed spectra has a value of about 2, which may be indicative of the physics of turbulence and raindrop coalescence and might suggest the predominance of a characteristic raindrop size. At 0.9 km resolution, the model rainfall spectra showed similar scaling to the observed down to about 10 min, below which a fall-off in variance was noted as compared to observations. Higher spatial resolution of up to 0.1 km was crucial to improve the ability of the model to resolve the shorter timescale variability. We suggest that the evaluation of dynamical models in the spectral domain is a crucial step in the validation of quantitative precipitation forecasts and assessing the minimal grid resolution necessary to capture rainfall variability for certain short timescales may be important for hydrological predictions.

DEVELOPMENT OF NEW DIAGNOSTIC TOOLS TO EVALUATE NWP CLOUD AND PRECIPITATION PRODUCTS USING A-TRAIN SATELLITE OBSERVATIONS

Paul A. Kucera, Courtney Weeks, Barbara Brown, and Randy Bullock

National Center for Atmospheric Research, Boulder, CO, USA,
e-mail: pkucera@ucar.edu

The goal of this presentation is to introduce new methods to evaluate NWP cloud and precipitation fields using high resolution A-Train observations (e.g., CloudSat, CALIPSO, etc.). For this effort, we are enhancing the NCAR Model Evaluation Tools (MET) to include the capability of ingesting satellite observations into a framework for comparisons with model products and/or other satellite datasets. The initial effort is focusing on the incorporation of CloudSat observations into MET for the comparison of NWP cloud products. Using CloudSat observations will allow for the unique opportunity vertical structure of cloud fields. In particular, we are developing the ability to match NWP products with satellite observations to provide diagnostic evaluation of attributes such as radar reflectivity structure, cloud top height, and cloud base height. We are currently applying these techniques to a variety of case studies ranging from large synoptic systems to tropical storms. The goal of this type of analysis is to provide tools to help improve the capabilities of models to forecast representative cloud and precipitation features. The presentation will give a summary of the new diagnostic evaluation capabilities being developed in MET. The presentation will also highlight some of the results obtained from analysis of several case studies.

CONDITIONAL RAINFALL VERIFICATION USING SYNOPTIC TYPING

Belinda Roux[1], *Alan Seed*[1], *Robert Dahni*[1], *and Ryan Smith*[2]

[1]*Bureau of Meteorology, Melbourne, Australia, e-mail:* b.roux@bom.gov.au
[2]*University of Manitoba, Manitoba, Canada*

It is commonly accepted that numerical weather prediction (NWP) models have varying skill in forecasting rainfall. There are numerous methods to verify model rainfall and the appropriate method depends ultimately on the forecast user. This verification of the NWP model rainfall is directed towards assisting hydrologists in using the data in hydrological models for flood forecasting and focus on the model's ability to reproduce rainfall in the right places and at certain thresholds, using scores like bias and correlation. In a preliminary investigation into model performance, eight case studies were performed over two areas in Australia, four in Victoria and four in south-east Queensland to account for the difference in latitude and weather systems (Roux and Seed, 2011). A clear difference was observed between the model performance in Queensland and Victoria, where the Victorian cases had a significant orographic component and were associated with large scale synoptic forcing, making them easier to forecast. From this study it seemed reasonable to conclude that the models perform better when the synoptic situation is organised on a large scale and has persisted for a number of days, and have problems in situations that are dynamic or organised on a small scale. But overall there was a wide spread in all the verification statistics which emphasises the point that NWP performance is highly variable, making it difficult to predict the accuracy of the predictions at the time that they are issued.

In an attempt to find the means to objectively forecast the predictability of an NWP model, a recent year of the latest model rainfall data is conditionally verified using the synoptic situation. The synoptic catalogue has been automatically derived using a synoptic classification application, Synoptic Typer Tools (Smith, 2008), and NCEP/NCAR Reanalysis data. From the verification results it should become clear how much confidence we can have in our NWP model rainfall predictions and under what situations.

Roux, B. and Seed, A. 2011. Assessment of the accuracy of the NWP forecasts for significant rainfall events at the scales needed for hydrological prediction

Smith, R. 2008. Synoptic Typer Tools: Designing and Testing an Automated Synoptic Classification Program. Hons. Thesis, Dept. of Geography, Uni. Of Winnipeg, Manitoba, Canada.

EVALUATION METHODS FOR TROPICAL CYCLONE FORECASTS

Barbara G. Brown

NCAR, Boulder, CO USA, e-mail: bgb@ucar.edu

Tropical cyclones are known worldwide to have some of the greatest impacts of all weather events, causing large amounts of damage from wind, extreme precipitation, storm surge flooding, and severe weather. In addition, tropical cyclones provide critical water resources in many areas of the world. Due to these large impacts, accurate prediction of tropical cyclones is a high priority for many weather services, and much effort has been extended toward improving these predictions. Traditional verification of tropical cyclone forecasts has focused on evaluations of the location of the storm center and the storm intensity as measured by the highest wind or lowest pressure associated with the storm. Many challenges are associated with verification of these basic predictions; for example, intensity values typically are inferred from satellite measurements since direct observations of the maximum wind speed are rarely available. Moreover, the "maximum" of any variable is typically very difficult to measure. As evaluations of additional variables (e.g., wind structure) are considered, and as new types of forecasts (e.g., ensembles) are implemented, new verification challenges are being faced. This talk will review current methods for evaluating the various attributes of tropical cyclone forecasts, consider challenges associated with application of these methods, and discuss new tropical cyclone verification approaches that are being developed.

RADIUS OF RELIABILITY: A DISTANCE METRIC FOR VERIFYING SPATIAL PROBABILITY FORECASTS

Elizabeth Ebert

Center for Australian Weather & Climate Research, Melbourne, Australia,
e-mail: e.ebert@bom.gov.au

High resolution forecasts of high impact weather such as heavy rain, severe thunderstorms, or tropical cyclone landfall, can have poor skill at point scale due to uncertainty in the precise location and/or timing of the event. To give credit to deterministic spatial forecasts that are close but not perfect, neighborhood verification methods such as the fractions skill score are now being used. These methods evaluate forecasts in the neighborhood of a point against the observation at that point, or to the corresponding neighborhood of observations. By verifying forecasts for a range of neighborhood sizes it is possible to determine the scale for which forecasts attain a certain level of skill.

A similar philosophy can be applied to probabilistic forecasts, to determine the spatial scale for which the forecast probabilities at a point are reliable. The existence of an observed event is assessed within a range of spatial neighborhoods centered on that forecast point. The neighborhood size for which the probability forecasts are reliable (over many forecasts) determines the "radius of reliability" (ROR). For a given forecast probability P, ROR can be interpreted as the search radius around the forecast within which at least one observed event can be expected to occur with frequency P. This concept is similar to the "cone of uncertainty" used in tropical cyclone track prediction to denote the region with 70% likelihood of enclosing the observed track. Similarly, the ROR can be used to convey the spatial precision of probabilistic forecasts. When used as a forecast quality metric, lower values of ROR indicate more skilful forecasts.

The ROR is best suited for evaluating and interpreting probabilistic nowcasts and short-range forecasts of rare events, such as those generated using convection resolving ensembles. The ROR will be demonstrated using forecast probabilities of heavy rainfall in landfalling tropical cyclones.

SOME NEW TECHNIQUES FOR HURRICANE VERIFICATION

Tressa L. Fowler, Eric Gilleland, and Barbara G. Brown

National Center for Atmospheric Research, Boulder Colorado, USA,
e-mail: tressa@ucar.edu

Verification information helps forecasters find models that improve on current hurricane forecasting capability. Many new metrics are being used and researched for assessment of both deterministic and ensemble tropical cyclone forecasts, some very simple and others more complex. Some multivariate analyses are proposed to examine the relationships between errors in different forecast fields.

For deterministic forecast models, average errors in track and intensity are examined by operational forecasters and verification systems. Some supplemental metrics have been introduced for recent assessments to examine both the frequency and magnitude of forecasting errors. Additionally, methods for making comparisons between a single model and several operational models have been added to improve the efficiency and understandability of verification results. These measures are simple and understandable, in the hopes of finding acceptance and use by the operational community. A rank histogram of errors compares a candidate model with a set of operational standards. Comparisons of two models are made using a frequency of superior performance for differences deemed practically significant. The traditional comparison of mean errors is enhanced by reporting p-values for the significance of the differences and making note of the practical significance of the typical error magnitudes.

Many new techniques for evaluating multivariate ensemble forecasts have been proposed recently (for example, minimum spanning tree rank histograms, energy score, determinant sharpness, etc.), but TC verification presents new challenges. In particular, track position is a quantity that involves both direction and distance errors. Great-circle distance is appropriate for determining the track error, but combining it with, for example, intensity error is more complicated. We are looking at the sensitivity of results to different methods of combining these disparate kinds of information. Additionally, we are exploring ways of diagnosing these errors simultaneously. For example, one method being investigated involves circle histograms that yield information about the distribution of wind intensity errors by along track errors.

Examples of these analyses and some proposed improvements for future assessments will be presented.

VALIDATION AND VERIFICATION METHODS APPLIED TO OCEAN FORECASTING

Gary Brassington

CAWCR, Bureau of Meteorology, Melbourne, Australia,
e-mail: g.brassington@bom.gov.au

Ocean forecasting has been established at a number of international centres serving a wide range of applications including: defence, maritime safety, marine accident and emergency, offshore engineering, and many others. The objective of ocean forecasting is to provide the best estimate and short-range forecast of the ocean state and circulation on the mesoscale. The mesoscale ocean is characterised by geostrophic turbulence which includes fronts, eddies (anticyclonic and cyclonic), barotropic/baroclinic instability, and eddy-eddy, eddy-mean flow and eddy-bathymetry interactions. Modelling and observing the mesoscale ocean is costly and challenging. The ocean has a Rossby radius of deformation on O(10-100) km, requiring high spatial resolution for both the ocean models and the observing system. The rapid attenuation of low frequency radiation also limits the observable properties from remote sensing to the near surface. The most important remotely sensed property is the sea surface height observed by altimetry. The in situ observing network has also been revolutionised through the introduction of Argo, a global network of autonomous floats, although the target coverage of 3 by 3 degrees is sparse relative the mesoscale.

Model validation and verification has been coordinated at an international level through the Global Ocean Data Assimilation Experiment and continues through the GODAE OceanView task team for Intercomparison and Validation. We will present an overview of how the intercomparison project has been organised and the future plans. In addition, there are validation efforts undertaken by each centre specific to their systems, particularly during the operational implementation cycle. Some examples will be presented based on the recent implementation of the Bureau of Meteorology's operational ocean forecasting system Ocean, Model, Analysis and Prediction System (OceanMAPS) version 2.

REGIONAL EVALUATION OF CONVECTIVE NOWCASTS AND APPLICATION OF CONTINGENCY TABLE PERFORMANCE DIAGRAMS

Barbara G. Brown[1], *Rita D. Roberts*[1], *Amanda R. S. Anderson*[1], *Eric Nelson*[1] *and Matthew Pocernich*[2]

[1]*NCAR, Boulder, CO USA, e-mail:* bgb@ucar.edu
[2]*Neptune and Co., Denver, CO, USA*

Evaluation of enhancements to convective nowcasting systems, and particularly nowcasts of convective initiation, is hampered by the impacts of large-scale storm characteristics as well as small-scale errors in timing and location. As part of an effort to evaluate the impacts of human forecasters on an automated convective nowcasting system (AutoNowCaster; ANC), some new approaches and evaluation tools have been applied, which provide remedies for some of these difficulties. This motivation for this work focused on an evaluation of a forecaster-interactive capability that was added to the ANC to allow forecasters to enhance the performance of one-hour nowcasts of convective storm initiation and evolution produced every six minutes. This Forecaster-Over-The-Loop (FOTL-ANC) system was tested at the U.S. National Weather Service Ft. Worth/Dallas Texas Forecast Office during daily operations from 2005-2010. The forecaster's role was to enter the locations of surface convergence boundaries into the ANC prior to dissemination of nowcasts to the Center Weather Service Unit. To identify the nowcast performance changes that might result from the forecasters' input, an evaluation approach was developed that aimed to isolate these changes from larger-scale performance variations. The approaches and methods developed for the evaluation include (a) focusing the evaluation on small regions rather than the broader domain of the ANC; (b) stratifying the cases into similar forcing categories; (c) examination of time series of performance for specific cases; and (d) use of contingency table performance diagrams to simultaneously investigate changes in multiple forecast performance attributes. These approaches together provide clear insights into the impact of the forecasters' input. The approaches also could have applicability to other types of forecast evaluation studies.

INCORPORATING TIMELINESS INFORMATION INTO AN ACCURACY MEASURE FOR CANADIAN WEATHER WARNINGS

Laurence Wilson[1], *Andrew Giles*[2], *and Erik Buhler*[3]

[1]*Meteorological Service of Canada, Montreal, Canada,*
e-mail: lawrence.wilson@ec.gc.ca
[2]*Meteorological Service of Canada, Edmonton, Canada*
[3]*Meteorological Service of Canada, Toronto, Canada*

Weather warnings are different from routine scheduled weather forecasts in that the forecaster decides both whether and when to issue the warning. If he tends to issue warnings early, with a long lead time, predictability considerations dictate that the warning is likely to be less accurate, obtaining a lower hit rate or a higher false alarm ratio or both. On the other hand, warnings that are issued with little lead time may be more accurate, but are likely to be less useful to the user community. The Meteorological Service of Canada has recently extended its regular forecast verification program to include weather warnings, and has requested that a measure be designed that incorporates both accuracy and timeliness information into a single score.

Two main characteristics of a good verification score are: 1. It is sensitive to the forecast attributes that are to be measured, and 2. It is "proper" in the general sense that it should not encourage the forecaster to develop a sub-optimal forecasting strategy simply to improve the score. In addition, the score should be as simple as possible so that it can be easily understood by non-specialist managers. In the present case, these two principle requirements mean that the penalty assigned to (too) short lead times should somehow balance the advantages of higher accuracy expected from short lead-time forecasts.

The new score, provisionally dubbed "Severe weather warning timeliness index" (SWWTI), is a modification of the familiar threat score to include an extra term which is a function of the lead time. In this presentation, the development of the score will be described, along with an analysis of its properties, based on the results of experiments using 2 years of weather warning data for selected Canadian regions.

ON USER, PROCESS AND MODEL ORIENTED WARNING VERIFICATION

Martin Göber

Deutscher Wetterdienst, Offenbach, Germany, e-mail: Martin.Goeber@dwd.de

The verification of warnings differs from model and point forecast verification mainly in the way warnings and observation are matched, since warnings are a kind of "block forecast" for a time and space interval. For the end users, warning verification should be based on events, i.e. for homogenised time intervals, since the user experiences a warning as a homogeneous event on which they have to react and the ups and downs of the weather during the interval are not relevant for the weather impact. The spatial homogenisation is governed by the main user, i.e. the emergency services, which operate on a county scale.

If the focus of the warning verification is on the warning process, than a more detailed e.g. hourly comparison can reveal other characteristics of the warning quality.

Another complicating factor which has to be taken into account is the variable lead time and duration of warnings. In order to accomplish a comparison to models an artificial "model like" sampling of warnings can be used.

All these approaches will be exemplified with operational thunderstorm and gust warnings from the German Weather Service DWD. It will be shown, that nowadays there are only few events which are completely missed by warnings, yet the strength of the event is sometimes underestimated. On the other hand, the number of false alarms has been drastically reduced since the introduction of county based warnings.

VERIFICATION OF OPERATIONAL (HIGH IMPACT) WEATHER WARNINGS

Tanja Weusthoff

MeteoSwiss, Zürich, Switzerland, e-mail: t.weusthoff@gmail.com

The increasing temporal and spatial resolution of numerical weather prediction (NWP) models requires advanced verification methods to reduce the so-called „double penalty" effect. Instead of a point-by-point comparison, new verification methods take into account the neighbourhood of the point of interest or compare objects in the forecast with objects in the observation.

Higher resolution NWP models also allow more precise weather warnings for small warning regions: MeteoSwiss, for example, issues now operational warnings for 152 regions in Switzerland. With the small regions, the problem of high-resolution model verification also becomes a topic for warning verification. Another important point is the dependency of warnings in space and time, which needs to be taken into account. This can be done by verifying consecutive warning events instead of hourly warning and observation pairs. MeteoSwiss is developing an automatic verification tool that allows a systematic and objective evaluation of the issued warnings. The method will be „tolerant" in terms of near hits and small displacements. Results from the verification of a 6 months test period in 2011 will be presented and the problems and challenges of an automated verification system for operational warnings will be discussed.

VERIFICATION OF WIND DIRECTION FORECASTS

Ian Jolliffe

University of Exeter, UK, e-mail: i.t.jolliffe@exeter.ac.uk

Forecasts come in numerous different forms, each of which may require different approaches to verification. One type of forecast for which verification is seldom discussed is wind direction. The forecasts could be categorical, typically one of 8 categories corresponding to compass points. A single category may be specified, or a probability distribution given over the categories. Alternatively, the forecasts may be continuous, given as angles measured clockwise from North. Verification for categorical and continuous forecasts has been well studied in the case where the categories or continuous variable can be linearly ordered, but the data here are different, lying on a circle. This difference necessitates adaptations to techniques developed for linearly ordered data. Some of these adaptations are described and illustrated.

QUANTIFYING VERIFICATION UNCERTAINTY BY REFERENCE DATA VARIATION

Manfred Dorninger and Theresa Gorgas

University of Vienna, Department of Meteorology and Geophysics, Vienna, Austria, e-mail: Manfred.Dorninger@univie.ac.at

The influence of reference data on NWP-model verification results is evaluated in the framework of a multi-level verification experiment. A set of six limited area NWP-models is evaluated by three model-independent analysis methods, based on two different observation networks. Verification is performed on four regular grids with horizontal resolutions ranging from 4-32km. In doing so traditional verification measures are combined with scale separation techniques using a 2-dimensional wavelet-transform. Verification uncertainties are estimated by four different applications: A poor man's ensemble derived from the sample of analysis variations, a re-sampling approach, and two different ensemble analysis tools based on random perturbations. The outcomes provide new insights in the interaction of grid resolution, observation density and spatial scale dependence of verification measures.

Variability of verification outcomes increases with larger grid point distance and lower observation density. In order to reduce uncertainty it is therefore advisable to apply gridded reference data on a high-resolution grid based on a preferably high-density observation network. A dense observation network lowers the importance of the interpolation methodology as it enables a simple interpolation technique to provide results equivalent to any advanced grid tool. The results indicate a clear preference of gridded reference data to pure observations. This is not only a matter of the station number which is considerably lower than the number of grid points in most analysis fields, but also a matter of the spatial representativity of observations. The verification experiments revealed that the question of scale dependence is even important for traditional, overall verification scores. If applied on field representations of individual spatial scales verification results for coarse scales show larger deviations or lower correspondence between forecast and analysis because of large amplitudes in the field. However, if verification is performed on ranges of scales, where only fine scale information is removed from the original fields, the scores provide improved results if compared to results of un-manipulated fields.

A CAUTIONARY TALE ON THE IMPACT OF OBSERVATIONS ATTRIBUTES ON FORECAST METRICS

Marion Mittermaier

Met Office, Exeter, UK, e-mail: marion.mittermaier@metoffice.gov.uk

It is generally agreed that observations are the preferred truth for verifying forecasts. Analyses may be gridded and provide a much greater data set for verification, but they are also generally dependent and constrain the forecast errors, painting a rosier picture of forecast systems than is actually the case.

Observations however are also considered "messy" and present several challenges in terms of their characteristics. Here we will provide two examples of how observation attributes and trends colour the verification results. One relates to long term trends in gridded radar-rainfall products, the other two manual and automated SYNOP cloud observations. Both show that if interpreted without an awareness of the characteristics and impacts, it could lead to drawing potentially false conclusions and taking the wrong set of actions.

POSTERS Day 1

VERIFICATION OF CCAM OVER JAKARTA REGION

Zainal Abidin, Indra Pratama and Eko Listiaji

BMKG, Public Weather Services, Indonesia, e-mail: zainal.abidin@bmkg.go.id

This verification focuses on skill of CCAM (from CSIRO) prediction related with daily precipitation and air temperature over Jakarta Region. The CCAM is the first NWP model that we run and it has been running operationally at BMKG since mid of 2010. We choose the daily precipitation because of the data avaibility limitation and the precipitation is the main meteorological parameter that directly related to social living in Indonesia. The air temperature parameter is from synoptic station around Jakarta Region, we chose this parameter to verify the CCAM as continous variable. The time range is only from August 2010 until July 2011. It is a preliminary verification that we have done, we used the simple method for verification, a dichotomus method for precipitation and RMSE for air temperature. The preliminary result shows CCAM tends to over-estimated.

UNDERSTANDING OF FORECAST IMPROVEMENTS IN JMA'S OPERATIONAL ONE-MONTH FORECAST SYSTEM DURING 2001-2010

Noriyuki Adachi, Yuhei Takaya, Masayuki Hirai, and Yuji Akasaka

Japan Meteorological Agency, Tokyo, Japan, e-mail: n-adachi@met.kishou.go.jp

The Japan Meteorological Agency (JMA) first started to run its operational one-month forecast system in 1996. Since then, the system has been improved in many aspects of a forecast model, data assimilation and ensemble methods. The system shows consistent improvements in forecast scores. Understanding and analyzing forecast improvements is crucial for continuous improvement in the future. To this end, we assessed the forecast performance of the operational real-time forecasts and hindcasts.

We verified traditional scores of Root Mean Square Errors (RMSEs) and Anomaly Correlation Coefficients (ACCs) with the real-time operational forecasts and latest hindcasts against JRA-25 reanalysis (Onogi et al. 2007) and JRA-25 real-time analysis (JCDAS), during 2001-2010. Supplemental 8-day forecasts with the JRA-25 system and deterministic high-resolution forecast systems are also analyzed.

Time series of operational one-month forecast scores for 28-day averaged 500-hPa height in Northern and Southern Hemispheres show continuous improvements for the last ten years (e.g., 5.11 m per decade for RMSE in Northern Hemisphere). It should be noted that the real-time operational forecasts show better performance than the hindcasts with the latest forecast model from around 2005 onward, if results of the same number of ensemble members (5 members) are verified for both real-time forecasts and hindcasts. The reason for this is as follows. The hindcasts were started from JRA-25/JCDAS initial conditions, which are analyzed with a 3-dimensional variational assimilation technique (3-D VAR). On the other hand, the real-time operational forecasts were started from the JMA's operational analysis, where 3-D VAR was replaced with a 4-dimensional variational assimilation technique (4-D VAR) in 2005. Therefore operational one-month forecasts made use of better analyses from the 4-D VAR. The 5-day forecast scores of the JMA's deterministic forecasts show better skill than the JRA-25 forecasts consistent with the one-month forecast result. The advantage of better initial conditions is also seen for week-2 forecast, suggesting that the quality of atmospheric analysis is important for extended-range forecast systems.

Detailed analysis will be presented including other impacts of changes in the JMA's one-month forecast system.

VERIFICATION AND OPTIMAL COMBINATION OF MULTI-MODEL PROBABILISTIC TERCILE SEASONAL FORECASTS

Trevor Carey-Smith[1]*, James Renwick*[1]*, Peter Thomson*[2]*, Brett Mullan*[1]*,*
Georgina Griffiths[1]*, Darren King*[1]*, and Andrew Lorrey*[1]

[1]*NIWA, Wellington, New Zealand, e-mail:* t.careysmith@niwa.co.nz
[2]*Statistics Research Associates Ltd, Wellington, New Zealand*

New Zealand's National Institute of Water and Atmosphere (NIWA) provides a seasonal outlook service to the general public consisting of probabilistic tercile forecasts of temperature, rainfall, river flow and soil moisture for six regions of the country (http://www.niwa.co.nz/our-science/climate/publications/all/seasonal-climate-outlook). This is currently a consensus forecast produced by expert assessment of a range of climate factors and seasonal climate models (e.g. NCEP; ECMWF). A comparison has been made of a selection of techniques (e.g. hit rate; Hanssen and Kuipers discriminant) to quantify the skill of NIWA's seasonal forecasts with reference to forecasts available from different global centres. In addition these verification techniques have been used to inform a method to optimally combine the available forecasts in an attempt to improve the current consensus approach. Methods to combine forecasts include a simple weighting based on overall verification metrics and more complex weighting to favour those models with greater recent skill or greater skill during the current ENSO phase.

WIND ENERGY PREDICTION AT NCAR'S RESEARCH APPLICATIONS LABORATORY

Daran L. Rife

National Center for Atmospheric Research, Boulder, Colorado, e-mail: drife@ucar.edu

This presentation highlights two wind power prediction capabilities developed at NCAR. The first is a new statistical postprocessing technique for making probabilistic forecasts, with an example application to near-surface predictions of wind, temperature, humidity, and pressure. The basic approach involves creating analog forecasts through the following two-step procedure. First, the current forecast is compared to all "past" forecasts of the same lead. Next, the dates of the closest pattern matches are noted. The matching patterns, called analogs, are ranked in terms of their similarity to the current forecast, and an ensemble is formed from the observed or analyzed conditions on those ranked dates. This new technique yields highly accurate forecasts that are strongly correlated with the observations. Attributes of the analog probabilistic predictions are estimated, including reliability, resolution, sharpness, and spread-error consistency. The presenter will show that the skill of the analog-ensemble is highly competitive with, and in some aspects superior to a state-of-the-science operational ensemble prediction system, the Environment Canada Regional Ensemble Prediction System (REPS).

The second part of the presentation describes the development of a more accurate and economical (less time, money and effort) wind resource assessment technique for the renewable energy industry, by incorporating innovative statistical techniques and new global mesoscale reanalyses. Initial results indicate that only a 1-2% sample of the 30-year record is needed. This means that the large computational cost of creating sub-kilometer downscaled weather simulations will be drastically reduced, but still produce a full wind characterization at a prospective wind farm.

COMPARATIVE SKILL ASSESSMENT OF CONSENSUS AND PHYSICALLY BASED TERCILE PROBABILITY SEASONAL PRECIPITATION FORECASTS FOR BRAZIL

Caio A. S. Coelho

Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), Instituto Nacional de Pesquisas Espaciais (INPE), Cachoeira Paulista, Brazil,
e-mail: caio.coelho@cptec.inpe.br

Tercile probability consensus seasonal (3-month mean) precipitation forecasts for Brazil for the categories below normal, normal and above normal precipitation are produced by CPTEC/INPE and displayed as spatial maps since early 2000. These forecasts are currently produced as a joint collaborative effort between CPTEC, the National Meteorological Service of Brazil (INMET) and regional meteorological centres of various regions in Brazil. The process for producing these forecasts consists in a) diagnosing the current global and regional weather and climate conditions recently observed, with particular emphasis on surface and subsurface ocean temperature conditions, the corresponding tropical convective activity and the associated high and low level atmospheric circulation response; b) examining the forecasts produced by physically based global and regional dynamical models and by empirical (statistical) models, including atmospheric-only models forced with forecast and persisted sea surface temperatures, coupled ocean-atmosphere models, regional models performing downscaling of the forecast produced by global models, and simple empirical models that use the most recently observed sea surface temperature conditions in the Pacific and Atlantic oceans as predictors for precipitation over Brazil for the next 3-month season; and c) use of climate expertise of all partners involved in this process together with the diagnostics and model forecast information previously examined to determined subjectively the consensus tercile probability forecast for Brazil.

Given the availability of a decade of consensus based forecasts for Brazil, and the current debate in the literature about the advantage these forecasts can provide when compared to physically based objective dynamical model forecasts, this study aims to perform a comparative skill assessment of tercile probability seasonal precipitation forecasts for Brazil produced by these two approaches. Standard verification scores typically used for assessing forecast quality of seasonal climate forecasts by comparing the forecast with the observed probability distribution function are used (e.g. Brier skill score, Ranked probability skill score). Such a comparative assessment aims to provide guidance on future practices for seasonal forecasting in Brazil.

TOWARD STANDARD PERFORMANCE METRICS FOR EVALUATING CLIMATE MODELS

Peter Gleckler[1], Elizabeth Ebert[2], Veronika Eyring[3], Pierre Friedlingstein[4], Helene Hewitt[5], Robert Pincus[6], and Karl Taylor[1]

[1]Program for Climate Model Diagnosis and Intercomparison, Livermore, USA, e-mail: gleckler1@llnl.gov
[2]Center for Australian Weather & Climate Research, Melbourne, Australia, e-mail: e.ebert@bom.gov.au
[3]Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany
[4]School of Geographic Sciences, Bristol University, Bristol, UK
[5]Met Office Hadley Centre, Exeter, UK
[6]Earth System Research Laboratory, Boulder, USA

The WCRP’s Working Group on Numerical Experimentation (WGNE) and Working Group on Coupled Modelling (WGCM) have jointly established a panel to identify a well‐defined set of performance metrics for climate models to objectively gauge the strengths and weaknesses of different models and to track improvement as models are further developed. The monitoring of climate model performance is intended to parallel similar effort by WMO to monitor the skill of numerical weather prediction models for short- and medium-range weather prediction, and dynamical coupled and statistical models for seasonal climate prediction. The panel brings together expertise in several key areas including global observations, forecast verification, process-oriented evaluation, and coupled model development and evaluation.

The panel is working to coordinate the development of a hierarchy of climate model performance metrics. At the most basic level will be a limited set of traditional “broad‐brush” statistical measures of large‐scale performance, such as bias, RMSE, and correlation of monthly averaged model fields against observations, evaluated over the globe. An extended set will focus on more targeted metrics to examine key processes and modes of variability (e.g., diurnal cycle, monsoon intensity, ENSO, carbon cycle, etc.). This will necessitate working closely with other international working groups and activities. The panel will encourage and facilitate performance metrics research by identifying key areas needing work and possibly organizing workshops.

Scientists at PCMDI will apply the standard performance metrics to relevant WCRP benchmark climate experiments included in the next phase of the Coupled Model Intercomparison Project, CMIP5. In particular, CMIP5 metrics will be computed and made public as the project advances. A "performance metrics package" will be made available to interested modeling groups and researchers.

DYNAMICAL INTRA-SEASONAL PREDICTION

Debra Hudson, Andrew Marshall, Guomin Wang, Oscar Alves and Yonghong Yin

Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, Melbourne, Australia, e-mail: D.Hudson@bom.gov.au

The Bureau of Meteorology has recently developed the capability to make dynamical intraseasonal climate forecasts (beyond the 10-days of a traditional weather forecast and shorter than a season), based on modifications to the coupled-model seasonal forecast system, POAMA (Predictive Ocean Atmosphere Model for Australia). The seasonal prediction version of POAMA was not designed for intraseasonal forecasting and has deficiencies in this regard. Most notably, the growth of the ensemble spread in the first month of the forecasts in the seasonal system, generated primarily from perturbed ocean initial conditions, is too slow to be useful on intraseasonal timescales, where forecast uncertainty is primarily determined by the atmosphere. This deficiency has been addressed through enhancements to the ensemble generation and initialisation strategy.

Through the application of appropriate verification methods, we will show that these modifications to the seasonal forecast system produce an ensemble prediction system which is appropriate for intraseasonal timescales. The forecast performance of the intraseasonal ensemble system is compared to the seasonal ensemble system over a common 20-year period of re-forecasts. Compared to the seasonal system, the ensemble spread of the intraseasonal system performs markedly better at tracking the root-mean-square error over the first month of the forecast. A direct consequence of this increase in ensemble spread is an increase in the reliability of forecasts (seen, for example, in forecasts of tropical SST and Australian climate) and a reduction in the sharpness of the forecast system (as seen by Attributes Diagrams). There are also improvements to forecast resolution (ROC Scores) and accuracy, but usually to a lesser degree. Interestingly, the verification also shows increased forecast skill on seasonal timescales (primarily the first season) in the intraseasonal system compared to the seasonal system. As such, the ensemble generation and initialization enhancements developed for the intraseasonal system are likely, in the future, to be incorporated into the seasonal system, producing a single, seamless intraseasonal to interannual forecast system.

COMPARISON OF CLIMATE CHANGE SCENARIO A1, A2, B1 AND B2 IN INDONESIA

Kadarsah[1]*, Ratna Satyaningsih*[2]*, Utoyo Ajie Linarka*[3]

[1]*Research and Development Center, The Indonesia Agency for Meteorology Climatology and Geophysics, Jakarta, Indonesia, e-mail:* kadarsah@yahoo.com

Climate change projection in Indonesia is a very interesting research topic. In a previous study, the Research and Development Centre of BMKG has conducted climate change research using A1B scenario of CSIRO Mk3.0 model. The results of climate change projections are very useful for climate change adaptation and mitigation in Indonesia. From several climate projections made by some international institutions, it is decided to use the CSIRO Mk3.0 model which is also used as an IPCC climate change model projection with horizontal resolution 20 x 20 km. The scenario being used is the A1, A2, B1 and B2 of the IPCC's scenario. Comparison of these scenarios is useful as a base of Indonesia suitable scenario. This leads to the adaptation and mitigation of climate change in Indonesia to be more focused and programmed. The study is using three time slices to indicate the period of changing, i.e., the present time (1979-2002), the near future (2015 – 2039) and the future projection (2075-2099). The projection results show that the climate change in Indonesia is significant mainly in the period of the end of the century. It is indicated by key parameters such as precipitation and temperature. The rainfall projection until the period of 2075 – 2099 shows an increase both in the periods of dry and wet seasons. In Java island for instance, the rainfall at the period of the dry season will be drier compared to the current situation.

ENSEMBLE ONE-KILOMETRE FORECASTS FOR THE SOUTH ESK HYDROLOGICAL SENSOR WEB

J. Katzfey and M. Thatcher

CSIRO Marine and Atmospheric Research, Aspendale, Australia,
email: jack.katzfey@csiro.au

An ensemble-based forecast system (60 km, 8 km, and 1 km resolution) using the Conformal Cubic Atmospheric Model (CCAM) has been set up on the Tasmanian Partnership for Advanced Computing (TPAC) cluster to predict rainfall over the South Esk catchment in Tasmania. The 60 km forecasts are for five days, while the 8 km and 1 km forecasts extend for three days. Ensemble members are based upon different initial conditions (using analyses generated at different national forecasting centres), different initial times, alternative downscaling approaches and varying the model's physical/dynamical configuration. This approach to ensemble forecasting is sometimes called a "poor man's" ensemble, although it has the advantage of being reasonably computationally efficient and can quickly identify a broad range of sources of forecast uncertainty. The ensemble forecasts are then used to provide for improved stream flow forecasts for the South Esk River.

The forecasts have now been running for more than a year and are being validated, both with gridded Australian Bureau of Meteorology analyses of rainfall observations and available station data. It is expected that the ensemble-based high-resolution forecasts will have a positive impact on the South Esk River stream flow predictions. Additional improvements in stream flow forecasting may result from merging of the model-based ensemble forecasts with available station data.

USING BIAS CORRECTION METHOD TO CALIBRATE FORECAST ENSEMBLES OF NMC

Li Li[1], *Li Yinglin*[1], *Tian Hua*[1], *Tao Shiwei*[1], *Cui Bo*[2]

[1]*National Meteorological Center, Beijing, China, e-mail:* lily@cma.gov.cn
[2]*National Centers for Environmental Prediction, Maryland, U.S.A*

Within the last decade, ensembles based on global models have been found useful for medium-range probabilistic forecasting. Ensemble forecasting has been embraced as a practical way of estimating the uncertainty of weather forecast and making probabilistic forecasts. However, ensemble forecasts still suffer from model and ensemble formation related shortcomings. These systematic errors would remain and cause bias in the 1st and 2nd moments of the ensemble distribution. In order to make a skillful medium-range forecast it is necessary to run post-processing algorithms to remove these systematic errors before the ensemble forecasts can be used. Thus, the main task of this study is to develop and implement a statistical post-processing scheme to reduce the biases in the National Meteorological Center ensemble forecasts against the data on analysis fields or observation fields. In this paper, we first investigate a statistical post-processing algorithm that is designed to adjust the 1st moment of ensemble forecasts.

The method introduced in this paper is an adaptive (Kalman Filter type) algorithm, which has the following application procedure: (a) Prior estimate to start up the procedure. At a given day T, calculate the time mean forecast errors between day T - 46 and T - 17 to initializes an average. (b) Update step. The average is updated by setting it to the weighted average of the new forecast error at day T - 16, with a weight of w, and the previous average, with a weight of $1-w$ ($0 \leq w < 1$). (c) Cycling: repeat step (b) every day from day T-15 to T-1. Such a decaying average bias assessment method is a convenient way to consider the most recent behavior of a system. The weight factor w controls how large the influence of the most recent data is. Experiments with different w (1%, 2% and 10%, respectively) have been conducted. In general, the 2% factor works better for most regions and seasons than the 1% and 10%.

This decaying average bias assessment method has been used to calibrate the bias in the 1st moment of the ensemble 500hPa geopotential height, 850hpa temperature and 2-meter temperature in NMC ensemble forecasts against the data on analysis fields or observation fields. After comparing the statistical variances such as absolute errors, anomaly correlation coefficient and Talagrand distribution, the conclusion could be drawn that the systematic errors of prognosis have been effectively decreased, and the ‘cold’ bias phenomenon of ensemble forecast also has been improved obviously. For two kinds of stations in which 2-meter temperature forecast is bigger and smaller than observation, the bias both has been calibrated towards observation. For the stations of which 2-meter temperature forecast is near to observation, the difference between raw and calibrated forecasts is not obvious, yet for most days, errors after calibration have been decreased compared to the raw forecast.

THE VERIFICATION OF REGIONAL ENSEMBLE PREDICTION SYSTEM PRECIPITATION FORECASTS AND ITS BIAS CORRECTION IN CHINA

Li Yinglin

National Meteorological Center of China Meteorological Administration, Beijing, China
e-mail: liyl@cma.gov.cn

The verification results of REPS precipitation against two different kinds of datasets respectively were compared, which one is observation on stations, and another is precipitation analysis products of CMA. The precipitation forecasts effect after bias correction were contrasted with before, too.

The results indicated that the scores calculated for the verifications against precipitation analysis have shown that the model performs better than the results against observations in many aspects, and obvious improvements have been seen in ensemble mean and probabilistic forecasts after the bias correction.

APPLICATION OF THE PATTERN MATCHING TECHNIQUE FOR CLIMATE CHANGE PROJECTIONS OF RAINFALL

Aurel Moise and Francois Delage

Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, Melbourne, Australia, e-mail: a.moise@bom.gov.au

Climate metrics are performance measures of global circulation models (GCMs) that have become one of the main foci within the climate model evaluation community since the publication of the IPCC Fourth Assessment Report. The range of suggested metrics varies from fairly simplistic ones – like the root mean square error – to the more elaborate ones using multi-model, multi-variable constructs. The World Climate Research Programme (WCRP) has recently initiated a panel with the task to identify and promote performance metrics for climate models. This study is addressing one of the aims of the WCRP panel, introducing a simple pattern matching technique applied to the mean state of the climate system. The technique is based on the object-orientated verification procedure CRA, and since then applied widely in the NWP community for spatial forecast verification of rainfall systems (for a detailed review, see Ebert and Gallus, 2009). Here we show an application of this technique by assessing the climate change projections of large scale rainfall structures such as the South Pacific Convergence Zone (Moise and Delage, 2011).

Ebert, E. E. and W. A. Gallus, 2009: Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification, *Wea. Forecasting*, 24 (5), 1401-1415.

Moise, A. F., and F. P. Delage, 2011: New climate model metrics based on object-orientated pattern matching of rainfall, *J. Geophys. Res.*, 116, D12108, doi:10.1029/2010JD015318.

EVALUATION OF THE ACCESS CLIMATE MODEL FOR THE CAPTIVATE PROJECT

Ian Watterson

CSIRO Marine and Atmospheric Research, Aspendale, Australia,
e-mail: ian.watterson@csiro.au

CAWCR's Earth System Modelling program has developed a new climate model ACCESS, based on coupling the Met Office's (MO) 'Unified' atmospheric model to GFDL's MOM4 ocean model. Evaluation has been an ongoing task during the development phase. The Model Evaluation group has contributed a range of assessments of climate. Skill scores for the basic fields indicate that ACCESS matches the best of the CMIP3 models for both the globe and Australia. The group has also joined the MO's 'CAPTIVATE' project, which aims to improve simulation of climate processes through further development of the HadGEM3 model. Our role is to perform evaluations for the Australian region, targeting important features of climatology, variability and teleconnections. A suite of simple objective tests is used, in order to allow multiple versions of models, including ACCESS, to be assessed. The results are summarized in a 'traffic light' form, allowing clear indication of shortcomings, improvements, and ultimately a 'fit for purpose' grading. In devising such tests, limiting data requirements and the suitability of observational data were important considerations.

Climatological features that are tested against observations include seasonal fields of surface temperature, pressure, and rainfall. Other features are the monsoon winds and east Australian subtropical jet. For each the non-dimensional 'M' skill score is used, and grades assigned. Variability and teleconnection testing has focused on the interannual variation of seasonal rainfall averaged over seven Australian regions used by the Bureau of Meteorology. Indices representing known drivers of this variability are tested. These include a high-latitude wind mode and two blocking indices, both derived from 500hPa winds. A single SST index for a Pacific-Indian dipole pattern is used. For variability, the standard deviation of each series is compared to that from observations. For teleconnection, the correlation coefficients between the rainfall and driver series are used. Results are averaged to provide an overall score for each model. In all these tests, ACCESS performs well, and matches the MO's models in most respects. The evaluations allow us to be confident that ACCESS will be a fine contribution to CMIP5 and a valuable tool for climate studies within CAWCR.

VERIFICATION OF TABULAR AIRPORT WEATHER BRIEFING

Peter Dunda

Bureau of Meteorology, Sydney, Australia, e-mail: p.dunda@bom.gov.au

The tabular Airport Weather Briefing (AWB), which is on trial at the Sydney Airport Meteorological Unit (SAMU), aims at enhancing the presentation of probability forecasts that are routinely provided within the text descriptions of an existing text AWB product. Feedback from users, particularly airline operators, suggested that the existing text format AWB could be improved by presenting the probability forecast information in a tabular format, which would be easier to incorporate into an end user's decision making process. The tabular format also enhances the ability to perform forecast verification and validation, which is difficult in the existing text AWB product.

A table of hourly probability forecasts covering critical weather elements and thresholds, which translate into impacts on air traffic operations, is provided as a supplement to the forecaster's text description of the weather, and to the ICAO stipulated aerodrome forecast. For example, the probability of fog or thunderstorm at the aerodrome, or the occurrence of low cloud ceiling and/or low visibility below specified operational thresholds, is given as a percentage at hourly intervals for the period covered by the aerodrome forecast. The tabular AWB enables the forecaster to provide additional levels of probability forecast information, especially when there is a small but realistic chance of critical weather elements reaching critical thresholds. Verification of the hourly probability forecasts provides valuable information on forecast performance that is difficult to measure in the text only product.

The probability forecasts can be verified on an hourly basis for each weather element and threshold to provide detailed information on forecast performance for various elements at different forecast lead times. Performance measures such as Probability of Detection (POD), False Alarm Ratio (FAR), the Relative Operating Characteristics (ROC) diagram or the Brier Score (BS) can be used to help forecasters identify performance targets and users to understand the value of forecasts in their decision making processes.

This paper shows some results of verification from the SAMU tabular AWB trial and discusses the advantages of producing forecasts and verification information in this format. This paper also aims to demonstrate the application of forecast verification to end user processes.

VERIFICATION OF NEW METEOROLOGICAL SERVICES IN THE TERMINAL AREA

Stéphanie Desbios[1], Bart Nicolai[2], Graham Rennie[3], Sandy Song[4], Jun Ryuzaki[5], Cecilia Miner[6], Cyndie Abelman[6], Steve Ricketts[7] and Peter Dunda[8] (for the WMO CAeM Expert Team on Meteorological Services in the Terminal Area, and Task Team / User Needs)

[1]Meteo France, Toulouse, France
[2]Belgocontrol, Steenokkerzeel, Belgium
[3]IATA/Qantas, Sydney, Australia
[4]Hong Kong Observatory, Hong Kong, China
[5]Japan Meteorological Agency, Tokyo, Japan, e-mail
[6]National Weather Service, Silver Spring, USA
[7]Meteorological Service of Canada, Edmonton, Canada
[89]Bureau of Meteorology, Sydney, Australia, e-mail: p.dunda@bom.gov.au

The new Meteorological Services in the Terminal Area (MSTA) initiative, which is being undertaken by the WMO in close collaboration with ICAO, aims at providing meteorological services to support Air Traffic Management (ATM) for wider terminal areas, especially those at busy airports, which are currently not covered by the standard meteorological services stipulated by ICAO. ICAO and the WMO are coordinating MSTA developments in advance of a projected endorsement in 2014.

MSTA forecasts aim to facilitate decision support. Hence, weather forecast elements would be translated into impacts on air traffic operations. These weather elements and the detailed thresholds/criteria to be used for the translation are to be determined as agreed between MET authority and user. Moreover, information on the level of confidence that the users can put on the forecasts is also required. This confidence level needs to be derived from verification and validation processes.

In terms of verification, the classic methods focus on measures such as mean absolute error or bias of continuously varying parameters such as temperature, pressure or wind (in direction and speed). These measures are difficult to incorporate into an end user's decision making process, unless they can be expressed in terms that are closely related to the user process, such as criteria or thresholds related to operating categories and procedures. Here the ability of the forecasting system to correctly predict the parameter in the given categories can be measured by performance indicators such as Probability of Detection (POD), False Alarm Ratio (FAR) or measures of success composed of these two. Evaluation of probabilistic forecasting systems is also possible using measures such as the Relative Operating Characteristics (ROC) diagram or the Brier Score (BS).

This paper shows some examples being used to verify forecasts which have impact to users. However, the use of POD/FAR to convey the accuracy of the forecasts could be problematic in some cases, especially where the weather impacts do not have clear thresholds/criteria. This paper discusses some of the issues involved in forecast verification and in communicating the results to downstream users.

AREAQNH FORECAST VERIFICATION METRICS

Nathan Faggian, Andrew Treloar, and John Bally

CAWCR, Bureau of Meteorology, Australia,
e-mail: nfaggian@bom.gov.au

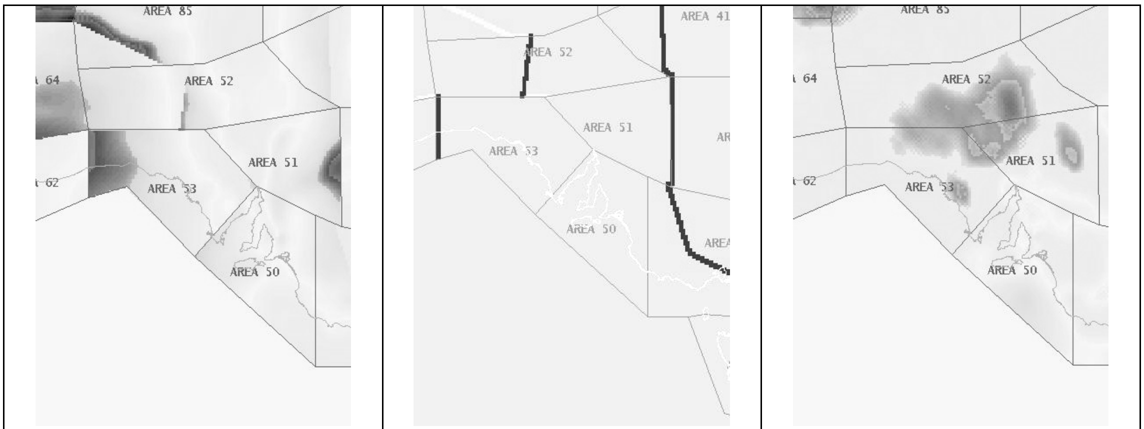
The Australian NexGenFWS has been recently configured to produce Aviation QNH (AreaQNH) forecasts for 31 zones across Australia every 3 hours. Area QNH is an atmospheric pressure value used to set aircraft altimeters and is critical for safe operations. It is subject to strict monitoring and amendment thresholds.

Aim: Provide metrics within NexGenFWS to verify AreaQNH forecasts with respect to the bias of numerical QNH guidance, AreaQNH forecast error and boundary consistency.

Methods: Using numerical guidance, a QNH field is derived (Treloar 2009) and AreaQNH values assigned to each forecast zone using a recursive value-assignment algorithm that attempts to minimize forecast error. Zones are sub-divided if threshold tolerances are exceeded. For verification purposes several error metrics are defined:

- Forecast and observations errors - the absolute difference between the AreaQNH forecast (hPa) and the forecast or analysed grid of QNH values
- Boundary error - the difference (using a sliding window) between adjacent AreaQNH forecasts
- QNH error - the difference between the numerical QNH guidance (hPa) and the analysed QNH values.

Results: Graphical representations of the error metrics are shown in Fig. 1. In practice, for introspection (or post-hoc analysis), each of the error fields are computed and archived. Notably, a test system has been in-place since late 2010 to provide automated computation of these metrics every 30 minutes.



a) **b)** **c)**

Fig. 1. Visual representation of the defined error metrics: **a)** AreaQNH forecast/observations errors, **b)** AreaQNH boundary consistency, **d)** QNH guidance versus QNH observations analysis.

Conclusion: Error metrics to track the quality and performance of AreaQNH forecasts are crucial to providing improved forecast processes in the NexGenFWS. Using the derived QNH, AreaQNH forecasts and up-to-date analysis, it is now possible to easily observe forecast quality over the entire forecast domain.

Treloar, A. 2009, “QNH Derivation and Forecasting in the GFE”, Bureau of Meteorology - *CAWCR Research Letters*, 3, 14-22.

SOME ISSUES WITH THE POINT DISTRIBUTION ON A CONTINGENCY TABLE FOR TAF PROBABILITY FORECAST VERIFICATION

Vivek C. Nagar

Australian Bureau of Meteorology, Melbourne, Australia, e-mail: v.nagar@bom.gov.au

Probability forecasts are frequently used within an aerodrome forecast (TAF) indicating the probability of occurrence of the weather events. These forecasts are used for the purpose of flight planning and for estimating the fuel requirement in the aviation operations. In Australia the commonly used probability types are 30% and 40% and are indicated as PROB30 and PROB40 on a TAF.

For the purpose of verification, it is slightly complicated to assign the proportionate points for a PROB30 or a PROB40 forecast on the contingency table. This is due to the fact that during a PROB period there are two forecasts running concurrently. The first forecast indicating the occurrence of the event, and the second forecast indicating the non-occurrence of the event. Ideally a forecaster should achieve best score when in a large sample the event occurs roughly the same percentage of times as given by the PROB statement. In other words when the probability of occurrence is stated to be 40, the forecaster should score best results when the event occurs 40% of times in a large sample.

In the present study it is shown that for a PROB40 forecast when a 0.4/0.6 point splitting scheme is used for occurrence/non-occurrence of the event, the best scores are achieved when a forecast verifies less that 2 out of 5 times, and the best score is recorded when the event does not verify at all. Similarly it is shown that in case of a PROB30 forecast when a 0.3/0.7 point splitting is used for occurrence/non occurrence of the event, the best score is achieved when event does not verify at all. These results hold good when the sample size is increased.

With a point splitting verification scheme in place for PROB forecasts, it is possible that the forecaster can develop a tendency of overly using probabilities whenever he/she has a low confidence in forecasting the event. There is another problem in verifying PROB forecasts using a point splitting method in a large sample. A forecaster can use the climatological probability of occurrence of an event every day and not try to add skill on top of it.

VERIFICATION OF FOG AND LOW CLOUD GUIDANCE FOR MAJOR AIRPORTS IN AUSTRALIA

G. T. Weymouth[1], T. Boneh[1], P. Newham[1], J. Bally[1], A. Nicholson[2], K. Korb[2], D. Sgarbossa[1], R. Potts[1], C. Matthews[1], C. Webb[1], and G. Elliott[1]

[1]Bureau of Meteorology, Melbourne, Australia, e-mail: G.Weymouth@bom.gov.au
[2]Monash University, Clayton, Australia

The Australian Bureau of Meteorology provides important fog and low cloud (FLC) forecasts for aviation. FLC forecasts are notoriously difficult, with uncertainty inherent in every component of fog forecasting. We developed a novel probabilistic approach for FLC forecasting, using Bayesian Networks (BN) that integrate uncertainty information. These networks probabilistically incorporate previous guidance and individual predictors, and produce suggested forecasts using decision theory. FLC networks were built for Sydney, Melbourne, Canberra and Perth Airports.

BN advantages include ease of forecaster understanding, the ability to provide objective probabilities for 'what if' scenarios, and easy, scientifically sound mathematical integration of subjective opinion and statistical forecasts, particularly when a predictor is unavailable or considered unrepresentative.

Verification metrics employed to verify the probabilistic forecasts include standard ROC scores and categorical statistics, as well as more unusual methods. We also compared guidance against official forecast skill (before and after guidance availability), previous guidance methods and climatology. Calibration of guidance has been undertaken.

In both operational trial and in cross-validation, BN forecast skill as measured by ROC area was better than the previous operational forecasts and guidance. At one stage. probability of detection of 100% was maintained for the operational guidance for 88 consecutive cases, with no increase in false alarm rate – 10 times the average previous number of events without a miss.

Observed fog frequencies were found to be higher when BN probabilities were large; however raw BN output in our networks was biased. Brier skill scores from calibrated BN output showed the BNs outperforming climatology.

Fog duration was found to be related to forecast fog probability, providing another verification of BN skill, and indicating the ability of BNs to provide information on fog duration. A confirmed corollary is that unforecast fogs are likely to be of shorter duration and impact than forecast fogs.

There is evidence that forecasters add value to the guidance in some situations.

BNs have proven most suitable for probabilistic forecasts of relatively rare events.

OPERATIONAL OPENROAD VERIFICATION

Robert Coulson and P. Gill

Met Office, Exeter, UK, e-mail: robert.coulson@metoffice.gov.uk

The automated operational OpenRoad verification is an integral part of the customer focused weather forecast package that helps road maintenance decision makers to plan, manage and minimise the effects of the winter weather.

The key benefits of verification are to validate the:

- accuracy and temporal nature of forecasts.
- forecasting skill using Kuipers' performance index, and critical success index.
- analysis of events in marginal conditions with economic values based on customer cost/loss ratios to evaluate cost-effective use of utilities.
- road state contingency tables to confirm confident decision making, regarding treatments to the network without compromising safety.

The results are visualised to enable clear communication to convey the effective measure and accuracy to decision makers as part of the verification tools.

THE VALIDATION OF DYNAMIC FIRE SPREAD MODELS: METHODS FOR COMPARING THE PERFORMANCE OF PERIMETER SPREAD PREDICTIONS

Thomas J. Duff[1], *Derek M. Chong*[1], *Peter G. Taylor*[2], *and Kevin G. Tolhurst*[3]

[1]*Dept. Forest and Ecosystem Science, Melbourne School of Land and Environment, The University of Melbourne, Burnley, Australia, email:* tjduff@unimelb.edu.au
[2]*Dept. Mathematics and Statistics, Faculty of Science, The University of Melbourne, Parkville, Australia*
[3]*Dept. Forest and Ecosystem Science, Melbourne School of Land and Environment, The University of Melbourne, Creswick, Australia*

Dynamic fire spread models are a recent development in fire management that provide for the simulation of fire spread through time in real landscapes under complex weather conditions. Fires in populated areas have the potential to cause catastrophic damage to life and property. Dynamic spread models allow for real time prediction and the evaluation of 'what-if?' scenarios, and can assist in tactical and strategic resource allocation. Such models are typically based on one dimensional spread models which are derived from empirical relationships. When considering spread predictions in three dimensions, there are few methods which can be used to objectively assess 'goodness of fit'. Effective objective methods of evaluating model performance provide for the discrimination of sources of error and can allow systematic improvement. We propose a number of new methods for assessing the performance of fire spread models. For a broad assessment of dissimilar perimeters, we propose the generation of 'pseudo-landmarks' on each perimeter to allow the use of landmark based analysis techniques. Both simulated and observed fires share a common origin and are typically driven by winds. These provide references for the creation analogous pseudo-landmarks, allowing for the assessment of indices such as orientation, size and residual perimeter difference. Where perimeters are highly congruent, we propose a method of estimating differences in fire travel path. The sampling of distances between simulated and observed fires on chords perpendicular to the perimeter of the simulated fire provides a spatially explicit indication of error. The location, direction and length of these chords can be used to determine input specific biases for model calibration and the isolation of error sources.

METOCEAN FORECAST VERIFICATION FOR OIL AND GAS EXPLORATION OVER SOUTH CHINA SEA

Ambun Dindang

The Malaysian Meteorological Department, Selangor, Malaysia,
e-mail: ambun@met.gov.my

This verification study is done for the Erb West site, an oil and gas exploration platform, located at latitude 6.3^{o} North and longitude 115.4^{o} East, approximately 83 km west-northwest of Kota Kinabalu, Sabah. The forecast data of significant wave heights (SWH), wind speeds (WSPD) and wave periods (WPRD) from the Malaysian Meteorological Department (MMD) Wave Model were used in this verification study. These forecast values are then compared against the observed values from the instruments at the platform. The bias and root-mean-square errors (rmse) for the T+24hr, T+48hr, T+72hr and T+96hr forecasts are then analyzed. The MMD Wave Model errors of SWH for the 24hr, 48hr, 72hr and 96hr forecasts are found to be small with maximum bias of 0.27 metres and minimum bias of -0.26 metres. The model also has low bias in predicting WSPD with maximum positive bias of 3.45 knots and minimum negative bias of -2.54 knots. The model has moderate bias and under forecast in predicting the WPRD with maximum positive bias of 0.37 seconds and minimum negative bias of -2.89 seconds. The rsme for SWH are found to be between 0.46 m and 0.16 m, WSPD between 5.86 knots and 3.39 knots, and WPRD between 2.82 second and 0.52 second.

ROUTINE COASTAL SEA LEVEL FORECASTS

Andy Taylor and Gary Brassington

CAWCR, Melbourne, Australia, e-mail: a.taylor@bom.gov.au

Routine coastal sea level forecasts have long been dominated by the established methods of tidal harmonic analysis and prediction. The associated institutions have evolved peculiar verification techniques and methods of communicating forecast uncertainty. In particular, uncertainties of tidal forecasts are almost never promulgated (if calculated at all) and quantitative verification is typically approached with regard to analysed amplitude and phase of tidal harmonics.

These peculiar practices are not fit for dealing with composite sea level forecasts made possible by the availability of operational oceanography products (OceanMAPS and real time tide gauge data). The verification requirements of these 'total sea level' forecasts pose some novel considerations. Additional aspects of verification methodology are raised by the newly implemented OceanMAPSv2.0 multi-cycle lagged ensemble.

This poster will outline the fundamental considerations of verifying total sea level forecasts and evaluate the merits of alternative perspectives; including continual time series, 'skew surge' and sea level extrema.

EVALUATE THE POTENTIAL ECONOMIC BENEFITS FROM WEATHER AND OCEAN RELATED INFORMATION TO BEACH USERS IN AUSTRALIAN EAST COASTAL AREAS

Fan Zhang and Xiao Hua Wang

School of Physical, Environmental and Mathematical Sciences, University of New South Wales at Australian Defence Force Academy, Canberra, Australia.
Fan.zhang@student.adfa.edu.au

The overall objective of this study is to estimate the potential economic benefits from improved weather and ocean related forecast information to beach users in Australian East Coastal areas. More specifically, it firstly attempted to learn the relationship between daily beach visitation densities in Gold Coast and Sydney and various weather and marine forecast information. The long term daily beach density changes corresponding to distinct weather and ocean forecast information are significant indicators to show visitors' demand of forecast information to plan their trips to the beaches in Australia. Secondly, the current forecast accuracy about weather and ocean information related to Gold Coast and Sydney areas will be examined to explain the forecast efficiency. Finally, the potential economic benefits from improved forecast information will be estimated through the simulations of beach visitor number changes to more accurate meteorological and oceanic forecast information. There are certain kinds of weather and ocean conditions that are preferred by most beach visitors, the tourists to Gold Coast and Sydney beaches with more accurate forecast information can make better decisions about the time of their trips to enjoy the beaches. On the one hand increase their enjoyment to beaches in good weather and ocean conditions and on the other hand decrease their trips under unpleasant weather and ocean conditions.

FORECAST VERIFICATION AS A MEANS OF STATISTICAL DECISION MAKING

Anatoly Muravyev

Hydrometeorological Research Centre of the Russian Federation, Moscow, Russia, e-mail: muravev@mecom.ru

Any decision, recommended or made, can be usually substantiated by means of skill scoring, error analysis and significance assessments, be these assessments statistical or heuristic by nature, and be the verification goals administrative, scientific, or economic.

The poster presents certain experience in forecast verification, accumulated at the Hydrometeorological Research Center of Russia, as well as 'decisions recommended or made'.

Examples of the long-range forecast activities embrace: (1) detailed spatial-temporal error analysis as feedback for modeler, (2) adaptive multi-criteria selection of hydrodynamic predictors, (3) ranking of forecast techniques on the bases of significance testing.

Short-range forecast verification problems and decisions to be made on the eve of the 2014 Winter Olympic Games include: (1) observation data network and reference archives, (2) COSMO-RU and ARW-WRF models, (3) domestic verification suites and international packages as VERSUS2, MET-MODE, SAL, CRA and Fuzzy.

EVALUATION OF NWP MODELS RESULTS OF WINTERTIME NOCTURNAL BOUNDARY-LAYER TEMPERATURES OVER EUROPE AND FINLAND

Evgeny Atlaskin[1,2] *and Timo Vihma*[2]

[1]*Helsinki University, Helsinki, Finland; e-mail:* evgeny.atlaskin@fmi.fi
[2]*Finnish Meteorological Institute, Helsinki, Finland*

Five operational numerical weather prediction (NWP) models were verified in winter conditions against (a) synoptic observations in Europe, (b) observations at a 48-m-high micrometeorological mast in Sodankylä, northern Finland, and (c) observations at the Helsinki Testbed stations (i) to evaluate the skills of the models to compute nocturnal 2-m air temperature (T2m) and the temperature inversion and (ii) to distinguish between the T2m bias and the subgrid-scale spatial variability of T2m. The results demonstrated a generally large positive T2m bias that increased with decreasing observed temperature and increasing inversion strength. When a strong temperature inversion was observed in Sodankylä, the models underestimated it, whereas in near-neutral conditions the stratification was overestimated. During coldest nights, the models strongly overestimated the initial T2m, revealing the problem also in data assimilation systems applied in the models. Validation of modelled T2m against grid-averaged T2m observed at Helsinki Testbed stations showed that the T2m model error dominated over the spatial variability of observed T2m. This suggests that over an almost flat terrain horizontal resolution is not a major factor for the accuracy of T2m forecast.

EVALUATION OF THERMODYNAMIC FIELDS FORECAST ACCURACY FOR DIFFERENT PHYSICAL SCHEMES IN THE WRF ARW MODEL

Vitalii Shpyg and Igor Budak

Ukrainian Hydrometeorological Institute, Kyiv, Ukraine, e-mail: Vilal@rambler.ru

In Ukrainian Hydrometeorological Institute the WRF ARW series models used from February of 2008. From that time to present these models recommended themselves as good and reliable modeling system which could be used for operational forecast and researches. In this study, we shown the results of numerical experiment which based on archive GFS data (NCEP, 1 degree of horizontal resolution) and carried out with the help of the WRF ARW v. 3.1.1 by using the combinations of five physical parameterization schemes. We conducted the numerical simulation for summer-autumn period (from 1-st July to 31-th of October 2009), two times (00 and 06 UTC) for each day with forecast length of 3 days. Horizontal step of domain was 30 km.

The spatial distribution RMSE, MAE and ME for temperature, pressure and wind velocity were analyzed for two times of model initialization. These simulated meteorological elements were verified against with 176 weather stations, which situated over all territory of Ukraine and provided observation 8 times per day. Spatially attention was attended for Carpathian Mountains region. For this area diurnal variations of meteorological elements above mentioned were considered.

We found that greatest values of forecast errors on the territory of Ukraine were occurred in mountains region (Carpathian and Crimea Mountains), across Black and Azov Seas coast and on the East of Ukraine. First result explained by established problem for mountains areas, second is connected with errors in sea-land mask for South of Ukraine, the third from them evidently has caused by river influence. The character of forecast errors changes with height was shown, it has clear relation between sea level pressure forecast errors and height and ambiguous for temperature and wind velocity forecast errors from one side and altitude from other side. The stations which have altitude more than 1000 m were not included in this analysis.

THE INTRODUCTION OF UNIFIED VERIFICATION AND COMPARISON PLAN FOR REGIONAL MODEL OVER CHINA AND ITS PRELIMINARY RESULTS

Chen Jing, Wang Yu, Li Li, Zhao Bin, Chen Fajin, Li Yinglin

CMA Numerical Prediction Center (CNPC), Beijing, China, e-mail: chenj@cma.gov.cn

The unified verification and comparison plan for regional model running in eight Regional Meteorological Center (RMC) of CMA, including Beijing, Shanghai, Guangzhou, Shenyang, Chengdu, Wuhan, Lanzhou, Urumqi respectively, was launched in 2006 and was operationally implementation in 2010. The models involved are GRAPES-meso, WRF, AREM, MM5. The flow chart of the data transferring and processing was established. The unified verification system was introduced. The verification variables are surface forecasts (including temperature at 2m, wind at 10m, precipitation) and high level winds, temperature and relative humidity at 850hPa, 500hPa and 250hPa, respectively. Verification parameters are TS, bias, hit rate, false alarm ratio and ETS for precipitation, and ME, MAE, RMSE, correlation coefficient, etc. for other surface and high levels variables. The daily verification results are issued on the web of NMC.

The evaluations of performance of regional models are performed in this study. The differences of verification parameters between the model of RMC and CNPC and their confidence intervals were studied. The results showed that the most RMC models had some advantages for the forecasts of precipitation but disadvantages for the forecasts of temperature at 2m. Moreover, the verification results are also able to provide some useful information for regional model developer. Finally the future plan for CNPC verification was discussed.

DEVELOPMENT AND VERIFICATION OF LDAPS

Yunho Park, Joohyung Son, Haejin Lee, and Hoon Park

Numerical Model Development Division, Korea Meteorological Administration, Republic of Korea,
e-mail: csat3@kma.go.kr

The KMA has set up a 1.5km resolution model (LDAPS: Local Data Assimilation and Prediction System) to predict high-impact weather phenomena in/around the Korean peninsula. LDAPS is a 3hr-cycling system with 3dvar and latent heat nudging. While both FGAT and IAU are used, RH nudging based on cloud fraction is not used in LDAPS. The system has three different variable grids (outside: 4km, middle: variable grid (4km to 1.5km), inside: 1.5km). The whole domain size is 744×928. The LDAPS is designed to produce 24 hour-forecasts eight times per day by assimilating data using 3dvar. We will show the configuration of LDAPS and several verification results with several case studies in a presentation.

PERFORMANCE ASSESSMENT OF THE OPERATIONAL ACCESS SYSTEM

Xiaoxi Wu

National Meteorological and Oceanographic Centre, Melbourne, Australia,
e-mail: x.wu@bom.gov.au

The Australian Bureau of Meteorology's Numerical Weather Prediction (NWP) system, the Australian Community Climate and Earth-System Simulator (ACCESS) replaced the Bureau's previous operational global, regional and tropical systems, GASP, LAPS and TXLAPS, in August 2010. The ACCESS system is a joint initiative led by the Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in cooperation with the university community of Australia. ACCESS is based on the Unified Model assimilation and prediction system at the United Kingdom Met Office (UKMO).

In this presentation, the operational ACCESS systems (i.e. global ACCESS-G, regional ACCESS-R and tropical ACCESS-T) are assessed by comparing with the GASP, LAPS and TXLAPS systems during the April to June 2010 quarter. ACCESS-G is also compared with GASP and those overseas global models operationally verified in NMOC (National Meteorological and Oceanographic Centre) during the same period. A similar comparison is also made for the period of April to June 2011 but without the discontinued GASP, LAPS and TXLAPS models. The fields of mean sea level pressure (MSLP) and 500hPa geopotential height have been selected for comparison. The operational overseas models used here include USAVM from the National Meteorological Center Washington, UKGC from UKMO, JMAGSM from the Japan Meteorological Agency and ECSP from the European Centre for Medium-range Weather Forecasting. The verification domain covers the Australian region. All the models are verified using a 2.5° latitude/longitude grid except USAVM which is verified on a 2.5° latitude/5.0° longitude grid - the use of coarser grid spacing for USAVM is not likely to affect the inter-comparison result greatly. The S1 skill score and anomaly correlation are the statistical methods employed here.

The results show the ACCESS system is much more skillful than the previous GASP and LAPS systems. Taking the commonly used cut-off of 60 per cent in anomaly correlation as the criterion for useful forecasts, ACCESS-G, ECSP and JMAGSM show useful skills to beyond 7 days, for the period from April to June 2011 ACCESS-G has a similar skill level to USAVM at the shorter lead times; JMAGSM is more skillful than ACCESS-G at 3 and 4 days, but becomes less skillful than ACCESS-G at day 6 and 7.

VERIFICATION OF NUMERICAL WEATHER PREDICTIONS OVER WESTERN SAHEL AFRICA

O.S. Idowua[1] and C.J. deW Rautenbachb[2]

[1]Department of Geosciences, University of Missouri, Kansas City MO, USA,
e-mail: oluseun.idowu@mail.umkc.edu
[2]Department of Geography, Geoinformatics and Meteorology,
University of Pretoria, South Africa

Processes in the atmosphere are not perfectly predictable in deterministic sense because the atmosphere is regarded as a non-linear dynamic and chaotic system. Numerical weather predictions (NWPs) are therefore subject to systematic errors and biases. The continuous verification of NWPs to improve forecast quality by better understanding forecast errors has become very important in recent years. This study verifies the 20 km by 20 km resolution Limited Area Model over Africa (Africa LAM) developed by the United Kingdom Meteorological Office (UK Met Office) to improve weather forecasts over western Sahel Africa (WS). The Africa LAM T+24h forecasts were verified against daily observed rainfall, maximum and minimum temperatures from 36 selected meteorological point stations across WS. Results indicate that the Africa LAM temperature forecasts show skill, more so during the raining seasons (April-May-June (AMJ) and July-August-September (JAS)) than during the dry seasons (January-February-March (JFM) and October-November-December (OND)) over the WS. The model rainfall forecasts, however, show more skill during the dry seasons (JFM and OND) than during the raining seasons (AMJ and JAS). The results further indicate that, on a regional basis, the model temperature forecasts show more spatial skill over the Wet Equatorial (WE) climate zone than over the Wet and Dry Tropics (WDT) and Semi Arid (SA) climate zones, while rainfall forecasts show more skill over the SA climate zone than over the WDT and WE climate zones.

VERIFICATION OF THE NEW 2.5 KM RESOLUTION LIMITED AREA MODELS AT CANADIAN WEATHER SERVICE

Francois Lemay

Environment Canada, Canada, e-mail: francois.lemay@ec.gc.ca

The Canadian Weather Service has recently made operational a new configuration of the GEM model at 2.5 km resolution based on what has been developed to provide accurate weather forecast for the last Vancouver Olympics in 2010. We present verification results that lead to this implementation and make comparisons with the previous 2.5 km configuration and also with our 15 km regional model. Issues related to different resolution model verification comparison are addressed. A calibration method of block bootstrap techniques to estimate score uncertainties is explored. Finally, we present a collaborative strategy to develop new verification tools for and with the help of users.

ON EVALUATING THE IMPROVEMENT OF THE NEW SIMM'S BOLAM USING A MULTI-METHOD APPROACH OVER THE MAP D-PHASE OPERATIONS PERIOD DATASET

Stefano Mariani and Marco Casaioli

Institute for Environmental Protection and Research (ISPRA), Rome, Italy,
e-mail: stefano.mariani@isprambiente.it

Since the beginning of the year 2000, it is operational at the Institute for Environmental Protection and Research (ISPRA, former APAT) an integrated meteo-marine forecasting chain called *Sistema Idro-Meteo-Mare* (SIMM). A cascade of four numerical models, telescoping from the Mediterranean basin to the Venice Lagoon, and initialized by means of ECMWF analyses and forecasts, forms the SIMM system. The forecasting system, and in particular the meteorological BOlogna Limited Area Model (BOLAM) at the base of the chain, has been regularly verified through subjective and objective methodologies.

In this framework, it has been decided to evaluate the quality of such new version and the achieved improvements with respect to the one previously implemented, referred as QBOLAM. In particular, this has been done using the rainfall observation dataset provided during MAP D-PHASE, a WMO-WWRP Demonstration project, when rain gauge measurements were collected from several regional and national networks available over the Alpine area and Central Western Europe. A reforecasting campaign has been then performed with the new version of BOLAM in order to provide forecasts over the D-PHASE Operations Period (DOP; Jun.-Nov. 2007), to be compared against the QBOLAM forecasts operationally delivered to the project during DOP.

The forecast verification has been performed by means of a multi-method approach based on a power spectrum analysis to assess difference in terms of (small) scale-details between BOLAM and QBOLAM forecasts, the computation of traditional categorical scores, a visual investigation on how contingency table's elements are spatially located over the verification domain, and the calculation of the ROC diagram. The performances of the two model versions have been also investigated by means of objected-oriented/spatial techniques with respect to two meteorological intense events occurred during DOP (25–28 Sep. 2007 & 22–25 Nov. 2007) over the Alpine area. Such investigation includes also the use of the precipitation fields retrieved from the radars available over the two target areas.

The work presents the improvements of the SIMM's BOLAM model in terms of quantitative precipitation forecast with respects to the previous implementation, focusing in particular on the results related to the two case studies.

ACTIVITIES OF UZHYDROMET

Fatima Kasimova

Centre of Hydrometeorological Service (Uzhydromet),
Tashkent, Republic of Uzbekistan,
*e-mail:*uzhymet@meteo.uz

Centre of Hydrometeorological Service (Uzhydromet) is the body of government administration which is specially mandated to fulfill the tasks in the field of hydrometeorology in Republic of Uzbekistan, and in its activities it is directly accountable to Cabinet of Ministers of RUz.
Strategic tasks of Uzhydromet:

- development and improvement of the government system of hydrometeorological observations;
- hydrometeorological provision of government and administrative authorities, fields of economy, armed forces and population of the republic with information about actual and expected conditions of environment and climate
- provision of urgent information about occurrence or threat of possibility of hazardous hydrometeorological phenomena;
- conduction of scientific-and-research works in the field of hydrometeorology, climate change and variation;
- improvement of short- and long-term weather forecasting, forecasting of river flow

Support of national objectives:

- improvement of quality of life of population
- enabling of high rates of sustainable development
- setup of capacity for the further development
- improvement of level of national security

Main operational and production subdivisions:

- "Meteoinfosystem" information-and-technical administration
- Administration for water inventory and meteorological measurements
- 13 territorial hydrometeorological administrations
- Service of monitoring of pollution of atmosphere, surface water and soil
- Service of meteorological provision for aviation
- Service of hydrometeorological provision

Organizations in-charge:

- Scientific research hydrometeorological institute
- Tashkent hydrometeorological professional college
- "Hydrometpribor" scientific-and-production enterprise

Climate service

- In the last years in our country as in the whole world the needs in climate service increased many times.

QUANTITATIVE PRECIPITATION FORECASTING FOR TWELVE AUSTRALIAN CITIES

James McGregor[1] *and Elizabeth Ebert*[2]

[1]*School of Geography, Environment and Earth Sciences, Victoria University of Wellington, New Zealand, e-mail:* jamesmcgregor@me.com
[2]*Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Australia*

This paper examines precipitation forecasts from the Australian Community Climate and Earth System Simulator (ACCESS), the American Global Forecast System (GFS) and the European Centre for Medium-range Weather Forecasts (ECMWF) weather models. It verifies model output against Australian Bureau of Meteorology rainfall analyses interpolated to twelve city locations throughout Australia. These cities are Adelaide, Alice Springs, Brisbane, Canberra, Darwin, Hobart, Melbourne, Perth, Port Hedland, Sydney, Townsville, Weipa. Up to two years of data are used covering the period from 1 July 2009 – 30 June 2011.

The chosen cities are representative of a variety of different climate regimes. Time-series plots of daily rainfall highlight these characteristics and also indicate significant inter-annual differences in rainfall at some of these city locations. The aim of this study is to use standard skill scores to compare and contrast model performance at each location and assess how performance may be related to climate type and also to measure how performance varies as a function of model forecast lead-time.

USING THE SEEPS SCORE TO ASSESS THE METUM AT THE UK MET OFFICE

Rachel North and Marion Mittermaier

Met Office, Exeter, UK, e-mail: rachel.north@metoffice.gov.uk

The Stable Equitable Error in Probability Space score (hereafter SEEPS) was introduced by Rodwell *et. al. (*2010). One of the strengths of the SEEPS is the use of climatological thresholds which enables climatologically diverse locations to be compared and aggregated. Thus far SEEPS has been used to assess daily precipitation totals, mostly from global models. For this study SEEPS has been used to investigate 6h forecast precipitation accumulations in the 12km resolution configuration of the Met Office Unified Model (MetUM). We will present initial results from analysing six-hourly precipitation accumulations over Europe originating from midnight forecast runs, and discuss the initial findings. These may include results showing the relative importance of climatological thresholds for shorter accumulation periods (24h vs. 6h). At present it is unclear whether the observed occurrence of, and regional models' ability (being higher resolution) to forecast short-period accumulation rainfall is potentially a function of the intensity only.

DID MELBOURNE FORECASTS IMPROVE AS A CONSEQUENCE OF GFE IMPLEMENTATION?

Harvey Stern[1] *and Benjamin Levin*[2]

[1]*Bureau of Meteorology, Melbourne, Australia, e-mail:* h.stern@bom.gov.au
[2]*Monash University, Melbourne, Australia*

Recently, there have been several changes to the forecasting process in the Victorian Office of the Bureau of Meteorology. Probably the greatest of these was the implementation of a system based upon the U.S. Gridded Forecast Editor (GFE) (in 2008), with its focus on point forecasts.

On the basis of official forecasts alone, it is difficult to establish whether or not the implementation resulted in more accurate forecasts. Ideally, one would need to have maintained a team of forecasters using the "old" methodology, and to compare that team's performance with that of the team using the "new" methodology.

We do not have two such teams. What we do have is an automated experimental forecasting system that, since 20 August 2005, has generated predictions in 'real time' by combining sets of official and other forecasts (this system is hereafter referred to as the "combined forecasts system" or CFS). CFS has operated essentially unchanged since implementation and it was therefore used as a "benchmark" to compare official forecasts with.

The skill displayed by predictions of four different weather elements (minimum and maximum temperature, amount and probability of precipitation) were calculated for the first year of CFS's operation (from 20 August 2005) and also for the sixth (from 20 August 2010). The measure of skill used was the correlation coefficient between predictions and observations, expressed as departures from normal.

During the first year, the average of the four correlation coefficients associated with official forecasts was 0.547 (CFS was able to improve these forecasts further such that the average was lifted to 0.609 – an increase of 0.062).

During the sixth year, the average associated with official forecasts was 0.601 (CFS lifted the average to 0.639 – an increase of 0.038).

	First year	Sixth year
Average correlation coefficient (Official)	0.547	0.601
Average correlation coefficient (CFS)	0.609	0.639

That the average associated with the official forecasts increased between year one and year six might be attributed to forecasts in the latter year being less difficult. However, that the CFS was able to increase the skill displayed by much less in the sixth year than in the first year suggests that the official forecasts have undergone a real improvement. That the largest recent change to how official forecasts are prepared was the implementation of GFE, it is concluded that they improved (at least in part) as a consequence of the GFE implementation.

VERIFICATION OF FORECASTER-PRODUCED GRIDDED FORECASTS

Philip Riley and *Chris Bridge*

Centre for Australian Weather and Climate Research, Melbourne, Australia,
e-mail: p.riley@bom.gov.au

The Australian Bureau of Meteorology Next Generation Forecast and Warning System (NexGenFWS), currently provides forecasts of several weather elements on a fine scale grid (3 to 6 km spatial resolution) for Southeastern Australia, and will be introduced to the remainder of the country over the next two to three years. Although the gridded forecasts are based on NWP guidance, the final version is produced by forecasters using a variety of grid-editing tools. Automated forecasts for several of the parameters are available from the Bureau's Gridded Optimal Consensus Forecasting (GOCF) system. Information on the skill of the resulting forecasts in comparison to what could be achieved using unedited GOCF guidance is important for such issues as editing policies and required staffing levels.

The Bureau of Meteorology has adopted the Verify software system created in the ECMWF for routine verification of gridded products, and this software is being adapted for the verification of the NexGenFWS forecaster-produced grids. This verification presents a number of challenges, for example: the forecaster produced grids are at a significantly higher resolution than the input NWP and GOCF guidance; forecasters perceive that they are correcting errors in the mesoscale analysis used as the verification standard, as well as in the guidance; and skill assessments need to account for the importance of the occasional correction of large guidance errors as well as the routine reduction of RMS error and bias. Early results of the verification of Temperature, Dewpoint and Precipitation forecast grids will be presented to illustrate how these challenges are being addressed.

VERIFICATION OF PRECIPITATION FORECAST - APPLICATION OF FUZZY METHODS FOR THE OPERATIONAL MODELS AT DWD

Ulrich Damrath

Deutscher Wetterdienst, Offenbach, Germany, e-mail: ulrich.damrath@dwd.de

There are two types of regional models at DWD that are running in an operational mode. COSMO-EU covers whole Europe and COSMO-DE is running for the region of Germany and some neighboured domains. The precipitation forecasts of both models are compared to radar information using fuzzy methods. Scores like fractions skill score and ETS for upscaling are calculated for model runs starting at 00, 03, 06, 09, 12, 15, 18 and 21 UTC, forecast intervals of three hours and different regions over Germany. The effect of competition of initialization using latent head nudging and the influence of boundary conditions is demonstrated both for cases during summer and autumn. The statistical significance of the results can be seen using the bootstrap technique. In order to show the convergence of bootstrap properties the effect of replication counts on the values of median and quantiles of ETS and FSS for different window sizes and precipitation amounts is also illustrated.

USING MODE TO TRACK THE SPATIAL CHARACTERISTICS OF CLOUD FIELDS

Marion Mittermaier

Met Office, Exeter, UK, e-mail: marion.mittermaier@metoffice.gov.uk

The UK weather is strongly affected by the advection of maritime boundary layer (BL) cloud over land, yet the upstream errors (over the sea) can not be assessed with conventional observations. Yet, if model developers are to improve the parameterisations that affect the development, persistence and break-up of these stratus or stratocumulus cloud sheets they need better diagnostic information. This is vital for improving temperature forecasts, especially in the winter months where the failure of one of these stratus cloud sheets to break up or advection errors lead to large temperature errors.

The evolution of cloud fields in the high-resolution Met Office Unified Model (MetUM) around the UK is assessed using the Method for Object-based Diagnostic Evaluation (MODE) and spatial cloud analyses or satellite-only cloud masks. In particular the method is applied to consider the onset, duration and size of cloud breaks. Here the way in which this is achieved is discussed. MODE's ability to provide useful statistics for cloud fields is considered, as they are quite different to precipitation.

USING IMAGE MAPPING TECHNIQUES TO GENERATE MODEL EVALUATION METRICS

Lawrie Rikus, Tracey Elliott and Gary Dietachmayer

The Centre for Australian Weather and Climate Research, Melbourne, Australia,
e-mail: l.rikus@bom.gov.au

In the past climate models have been evaluated using a number of qualitative processes, including visual inspection of plots of model fields and subsequent expert judgment. As part of a system to produce a system for targeted objective assessment of climate models a pattern identification program (PyBlob) is been developed to automate the assessment of patterns and produce quantitative results suitable for inclusion in model metric sets. To provide an illustration and an initial test of the technique, the program has been applied in a study of the atmospheric jets in a number of observational datasets. PyBlob has been used to quantify the properties of the sub-tropical jets defined as closed contour levels on monthly zonal mean zonal wind plots. The resultant time series have been examined using a number of statistical techniques to attempt to assess the uncertainty in the observations. The extension to climate model metrics will be discussed as well as a number of other applications and extensions for PyBlob for model evaluation in climate modelling as well as numerical weather prediction.

APPLICATION OF SPATIAL-TEMPORAL FRACTIONS SKILL SCORE TO HIGH-RESOLUTION ENSEMBLE FORECAST VERIFICATION

Le Duc[1]*, Kazuo Saito*[2]*, and Hiromu Seko*[2]

[1]*JAMSTEC, Yokohama, Japan, email:* leduc@jamstec.go.jp
[2]*MRI, Tsukuba, Japan*

Two ensemble systems with the resolutions of 10 km (MF10km) and 2 km (MF2km) were compared to verify the value of cloud-resolving ensemble forecast in predicting 1-hour precipitation, using Fractions Skill Score (FSS) as a measurement. FSS as defined is applied for deterministic precipitation forecasts with spatial neighborhoods. By recognizing both spatial mismatch and timing lag problems inherently in short period precipitation forecasts, the concept of spatial neighborhood of FSS was extended to temporal one. Instead of a neighbor area in space, a neighborhood should be rather understood as a spatial-temporal window. In addition to this, FSS idea was also applied for ensemble forecasts by including all members into each neighborhood when each member from an ensemble forecast can be considered as a possible realization of the true state. This is somehow similar to adding a new dimension to spatial-temporal neighborhoods.

The verification results showed that MF2km was better than MF10km in heavy rainfall forecasts significantly. In contrast, lower-resolution forecasts MF10km beat high-resolution forecasts MF2km for light rains, suggesting that the horizontal resolution of 2 km is not necessarily fine enough to completely remove the convective parameterization. For moderate rains, two systems exhibited similar performances where the FSS differences were not statistically significant.

NEIGHBOURHOOD VERIFICATION IN SPACE AND TIME

Tanja Weusthoff

MeteoSwiss, Zürich, Switzerland, e-mail: t.weusthoff@gmail.com

The operational precipitation forecasts of the Swiss numerical weather prediction (NWP) models COSMO-2 and COSMO-7 are evaluated with reference to Swiss radar measurements by means of neighbourhood verification methods (Ebert, 2008). Based on earlier performed idealized case studies, two different methods were chosen for the identification of an event: (i) the mean value over the respective window exceeding a given threshold (Upscaling) and (ii) the fraction of grid points exceeding the threshold (Fractions Skill Score), respectively. The verification is done on different spatial scales by varying the window sizes and for various thresholds defining an event. The scores are determined for 3h rain accumulations and are then aggregated on different time scales.

The here presented verification of precipitation forecasts focuses on the comparison of the two operational COSMO models at MeteoSwiss with a resolution of 2.2 and 6.6 km. By varying the spatial scales the scores are calculated for comparable scales allowing for a direct comparison of the models' skill. The robustness of the results is determined by means of a bootstrapping procedure. In addition, the dependency of the models performance on the weather type, classified in an objective way by means of the new automated weather type classification introduced at MeteoSwiss is presented.

In order to add a tolerance in time, the spatial window of the neighbourhood verification has been extended by a temporal window ("fuzzy in time").

MOS AND THE 2X2 CONFUSION MATRIX

Klaus Knüpffer[1] and Stefan Metzner[2]

[1]*METEO SERVICE weather research GmbH, Berlin, Germany, e-mail:* klaus@mswr.de
[2]*Free University Berlin, Institute of Meteorology, Berlin, Germany*

A developer of MOS (Model Output Statistics) systems can be faced with challenging verification situations. One of them is a statement that MOS Yes/No fog forecasts for the upcoming hour perform worse than the persistency of the latest observation - which is part of the MOS predictor set. A closer examination of this scenario has resulted in a variety of empirical studies, the results of which shall be presented.

An initial analysis showed that MOS performed indeed worse if categorical visibility forecasts are translated into Yes/No decisions for fog and if the measure of performance is the TSS (True Skill Statistic). However, alternative performance measures like HSS (Heidke Skill Score) and PC (Percent Correct) showed advantages of MOS w.r.t. persistency in this forecast situation.

Motivated from these contradicting verification results, a series of empirical studies investigating the dependence of the above-mentioned and other verification scores to an artificially introduced forecast bias was conducted. For each performance measure, requiring either under-forecasting (e.g. PC), bias-free forecasts (e.g. HSS) or over-forecasting (e.g. TSS) of a rare event, various scenarios are provided. The extent of over- and under-forecasting depends on two parameters: The event probability and the predictability. A MOS developer can quantify not only the former but also the latter.

Optimum thresholds for the conversion of probabilistic forecasts into categorical forecasts are provided w.r.t. these and other performance measures. For PC this threshold is 50% while for TSS it is the event probability and for HSS the threshold is a function of event probability and predictability. Thus, the current research provides guidelines on calibrating MOS forecasts so that they deliver optimum Yes/No decisions depending on the verification score applied.

These findings are emphasising the limitations of TSS as a performance measure for (rare) event forecasts. It is concluded that considering the high incidence of false alarms, forecasters should not disseminate TSS optimised forecasts for rare events to the general public.

FROM READING TO MELBOURNE: VERIFICATION WITH PYTHON

Chris Bridge[1], *Beth Ebert*[1], *Xiaoxi Wu*[1], *Claude Gibert*[2]

[1]*Bureau of Meteorology, Melbourne, Australia, e-mail:* c.bridge@bom.gov.au
[2]*Synoptic View, United Kingdom, e-mail:claude@synopticview.co.uk*

'Verify' is a gridded verification tool. A Python-based software package, it was initially developed at the European Centre for Medium Range Forecasting (ECMWF) at Reading in the UK.

The ECMWF kindly provided Verify to the Bureau of Meteorology here in Melbourne where it has been adapted as a general gridded verification framework available for both operational and research use across the organisation.

We describe some of the successes and challenges faced during a project that began as a 6 month pilot and is now well into its third year.

THE DEVELOPMENT OF GRAPES EVALUATION TOOLS (GET)

Zhao Bin and *Chen Jing*

National Meteorological Center, China Meteorological Administration, Beijing, China,
e-mail: zhaob@cma.gov.cn

Based on the operational statistics and diagnosis methods of national meteorological center, a new verification system for GRAPES_GFS named GET (GRAPES EVALUATION TOOLS) has been developed. The statistics methods such as the ACC and RMSE and the significant test at 95% confidence level are presented. The diagnosis methods of forecast bias and standard deviation are also calculated within.

It displays all verification statistics for global NWP forecasts and data used on the binary as the GRAPES model output format and the regions referred to are: Global (GLOB), NHEM (20°N-90°N), SHEM (20°S-90°S), TROP (20°S-20°N), EASI (70°E-145°E, 15°N-65°N). Pattern correlations are computed using anomalies respective to climatology of the NCEP/NCAR reanalysis. It also carried out a variety of diagnosis methods for testing different physics and dynamics packages of the model.

Using a frame of modules, the evaluation tool is worked with shell script for evaluating the GRAPES_GFS forecasts and for comparing the different simulation results with other GRAPES tests. Six simulation results are permitted to compare with GET, and the evaluation results generated as JPG format can show with a web page. The web site provides a simple mode for all researchers to achieving important simulation comparison information.

RECENT ENHANCEMENTS TO THE MODEL EVALUATION TOOLS (MET4.0)

Tressa L. Fowler, John Halley Gotway, Randy Bullock, Paul Oldenburg, and Tara Jensen

National Center for Atmospheric Research, Boulder Colorado, USA,
e-mail: tressa@ucar.edu

Model Evaluation Tools (MET) is a freely-available software package for forecast verification. It is distributed through the Developmental Testbed Center (DTC) for testing and evaluation of the Weather Research and Forecasting (WRF) model. Development has been led by the community: including WRF users, the DTC, and verification experts through workshops and user meetings. MET allows users to verify forecasts via traditional, neighborhood, and object-based methods. To account for the uncertainty associated with these measures, methods for estimating confidence intervals for the verification statistics are an integral part of MET.

Many features of the basic software were presented at the last workshop. The latest release includes many new features. The new Ensemble-Stat tool preprocesses sets of forecasts into ensemble forecasts, including mean, spread, and probability. When observations are included, it will also derive ensemble statistics such as rank histogram and continuous ranked probability score. When accumulating statistics over time, users can now adjust the confidence intervals to account for serial correlation. To assist our WRF ARW users, MET now can read the netCDF output from the pinterp postprocessor. Multi-category, e.g. 3x3, contingency tables are supported, along with appropriate skill scores. Many new preprocessing tools assist users with formatting and examining their observational data for use with MET. The tools reformat MADIS (Meteorological Assimilation Data Ingest System), TRMM (Tropical Rainfall Measuring Mission) and WWMCA (World Wide Merged Cloud Analysis) products and produce plots of many types of observational data sets. These enhancements will be of particular interest for users performing cloud forecast verification. Examples of the existing and new verification capabilities will be shown.

ADVANCED VERIFICATION ACTIVITIES IN COSMO CONSORTIUM

Ulrich Damrath[1], Flora Gofa[2], Elena Oberto[3], Adriano Raspanti[4], Francis Schubiger[5], Vanessa Stauch[5], Maria Stefania Tesini[6], Tanja Weusthoff[5]

[1]DWD, Offenbach,Germany, Ulrich.Damrath@dwd.de
[2]HNMS, Athens,Greece, fgofa@hnms.gr
[3]Arpa-Piemonte, Torino, Italy
[4]USAM-CNMCA,Roma, Italy
[5]MCH, Zurich, Switzerland
[6]ARPA-EMR,Bologna, Italy

Verification activities are an essential part of numerical weather predictions, crucial for the development, refinement and interpretation of any modeling system. Usually a comparison between model predictions and observations provide useful information on quality or accuracy, but the results depend on the methods and metrics used to verify as well as on the different aggregation of both forecasts and/or observations, taking into account the increasing resolution of numerical models. Moreover the interpretation of the huge number of plots produced is far from being trivial and even more complicated is the evaluation of model deficiencies and drawbacks. An effort to improve the use of verification has been undertaken in COSMO consortium in these last years through the use of the COSMO unified system for verification (VERSUS) and also through new methods of representation of results. Particularly in this study it will be shown, the use of conditional verification as a tool to explore interdependencies between various weather parameters values and model errors, the use of performance diagrams as a way to compare with a glance the performance of various models ("father and son"), the potentiality of fuzzy verification and the important indications that the study of long-term trend errors can give. Furthermore in these last two years, a special effort has been done to create a link between verification and forecast activities, trying to identify weather situation-dependent weaknesses and strengths of COSMO model, with the use of a stratification based on certain weather regimes.

VALIDATION OF TRMM DAILY PRECIPITATION ESTIMATES OF TC RAINFALL USING PACRAIN DATA

Yingjun Chen[1], *Elizabeth E. Ebert*[2], *Kevin Walsh*[1] *and Noel Davidson*[2]

[1]*University of Melbourne, Melbourne, Australia, e-mail:* y.chen@bom.gov.au
[2]*Centre for Australian Weather and Climate Research, Melbourne, Australia*

Since its launch in 1997, Tropical Rainfall Measuring Mission (TRMM) precipitation products have been widely used in tropical cyclone (TC) rainfall studies. As satellite data are indirect estimates of rainfall, it is important to evaluate their accuracy and expected error characteristics. Previous validation studies were mainly made over land, with no TRMM validation studies of TC rainfall having been done over the ocean before.

This study is to validate TRMM daily precipitation estimates of TC rainfall using the Comprehensive Pacific Rainfall Database (PACRAIN) of 24h rain gauge observations. The evaluation is done both on atoll sites, which represent open ocean, and coastal sites and islands, which may be influenced by the topography, so that we can compare TRMM's performances for different terrain. The results show that TRMM 3B42 has reasonable skill at detecting TC rains. The performance on atoll sites (over ocean) is better than on coastal sites (over land), and it is better able to estimate the intensity of the TC heavy rains on atoll sites. This gives users more confidence to use TRMM data to analyse TC rainfall over the ocean. TRMM 3B42 is less impressive at coastal sites where it underestimates TC heavy rains, suggesting that TRMM is not able to detect orographic enhancement during TC landfall.

VERIFICATION AND APPLICATION FOR 500HPA HEIGHT FORECASTS OF TWO NWP MODELS DURING THE PERIOD OF TYPHOON MUIFA USING OPTICAL FLOW TECHNIQUE

Zhihui Zhu and Ningli Huang

Shanghai Marine Meteorological Center, Shanghai, China, e-mail: zzh19841116@yahoo.cn

Using optical flow technique, 500hPa height forecasts of two numerical weather prediction (NWP) models ECMWF (EC for short) and T639 during the period of Typhoon MUIFA are verified and interpreted. The optical flow technique is found to yield useful information on the distortion error over the typhoon region in the forecasts. EC has stable errors in 24-h, 48-h and 72-h forecasts. Its 24-h forecast intensities are 0-10gpm lower than observations, and its 24-h forecast angular errors are between 180 and 240°, which are between west and southwest. Its 48-h forecast intensities are 10-20gpm lower than observations, and its 48-h forecast angular errors are also between 180 and 240°. Its 72-h forecast intensities are mainly 20-30gpm lower than observations, and its 72-h forecast angular errors are mainly between 180 and 270°. As a contrast, the intensity and angular errors of T639 are not so stable. As for displacement errors, the performances of EC and T639 are both unstable. The similar disparities between observation fields and 24-h, 48-h and 72-h forecast fields of the two models also show that EC has better forecasting performances than T639. Finally, an application technique is used to correct the 48-h forecast fields according to the 24-h forecasts errors. The results show that when the distortion errors of 24-h forecasts obtained from optical flow field are used to correct the 48-h forecasts, the correction fields are more similar to the observation fields than the forecast fields itself.

THE VERIFICATION OF TROPICAL CYCLONE RAINFALL PREDICTIONS WITH GLOBAL MODELS OF CMA AND JMA IN 2010

Wang Yu

Numerical Weather Prediction Center,
China Meteorological Administration,Beijing, China,
e-mail: wangyu@cma.gov.cn

This paper verifies the performance of numerical prediction on landfalling tropical cyclone (TC) associated precipitation in China in 2010, the year when the landfalling TCs over mainland China and their associated rainfall are less than the average. The verification method we used in this work is the one operationally used by National Meteorological Centre (NMC) / China Meteorological Administration (CMA). The verification results over Guangdong, Guangxi, Fujian and Hainan provinces, where TC rainfall is relatively severe comparing to other regions, indicate that model forecasts for Guangdong and Guangxi provinces are more skillful than those for Hainan and Fujian provinces.

The performance of TC rainfall forecast by the global models of CMA and Japan Meteorological Agency (JMA) are compared. Although with different schemes of physical processes, data assimilation and dynamic framework, the two models perform similarly in TC rainfall and track forecasting for all TCs. The rain belts predicted by the two models within 60 hours are in good agreement with observation in terms of both the location and magnitude of maximum rainfall. Beyond 3 days, the predicted rain belt shifts more to the east and is less in amount than the observations. JMA model outperforms CMA model in the forecast of TC rainfall location, the latter tends to false alarm heavy rainfall processes.

ENSEMBLE FORECAST VERIFICATION FOR TROPICAL CYCLONES

David Smith, Michael Naughton and Asri Sulaiman

Bureau of Meteorology CAWCR, Melbourne, Australia, e-mail: dhsmith@bom.gov.au

The ACCESS global and regional ensemble prediction system (AGREPS) is now running daily, producing 5 day global and 3 day regional forecasts, with 24 members. This poster will review the ensemble's basic characteristics, and its future planned development. After examining some current forecast results in a general context, the focus will shift to recent extreme weather events in the form of tropical cyclones. Cyclone tracks and calculated verification results will be presented, in both standard and multi-scale formats, the latter involving a multi-scale decomposition tool applicable to local domains.

VERIFICATION OF EXPERIMENTAL MODELS FOR TROPICAL CYCLONE FORECASTING IN SUPPORT OF THE NOAA HURRICANE FORECAST IMPROVEMENT PROJECT (HFIP)

Paul A. Kucera, Barb G. Brown, Louisa Nance, and Christopher L. Williams

National Center for Atmospheric Research, Boulder, CO, USA, e-mail: pkucera@ucar.edu

In 2009, the National Center for Atmospheric Science (NCAR)/Research Applications Laboratory's (RALs) Joint Numerical Testbed (JNT) Program formed a new entity called the Tropical Cyclone Modeling Team (TCMT). We have been developing new methods and conducting diagnostic verification experimental of tropical cyclone forecast models. For the Hurricane Forecast Improvement Project, the TCMT focuses is on designing evaluation methods and conducting verification analysis for various regional and global forecast models that participate in the annual HFIP retrospective and near real time forecast demonstration studies. The TCMT is also developing statistical approaches that are appropriate for evaluating a variety of tropical cyclone forecast attributes. These methods include new diagnostic tools to aid, for example, in the evaluation of track and intensity errors, ensemble forecasts, and tropical cyclone genesis. During 2011, the TCMT conducted a retrospective evaluation of eight experimental models to determine if they were possible candidates for real-time evaluation for the 2011 northern Atlantic and Eastern Pacific Ocean hurricane season. This retrospective analysis was conducted using a subset storms selected by NHC from the 2008-2010 hurricane seasons. An overview of the new methods being developed along with a summary of results from the 2011 retrospective analysis will be provided the presentation.

ESTABLISHMENT AND QUALITY EVALUATION ON MESOSCALE METEOROLOGICAL ANALYSIS FIELD IN SOUTH CHINA HEAVY RAINFALL EXPERIMENT

Hongli Li, Juxiang Peng and Minghuan Wang

Wuhan Institute of Heavy Rain, CMA, Wuhan, China, e-mail:
maxiyouyou@yahoo.com.cn

In this study, the observational data acquired in the South China Heavy Rainfall Experiment from May to July, 2008 and 2009 were integrated and assimilated through an application of US-made LAPS system introduced by Wuhan Institute of Heavy Rain. Then, a mesoscale meteorological analysis field was generated with a spatial resolution of 5km and a temporal resolution of 3 hours in four observational areas of Southern China, Central China, Jianghuai and Yangtze River Delta.

An evaluation analysis was carried out on the quality of LAPS mesoscale analysis field in two aspects. First, aiming at the key heavy rainfall events, LAPS meso-scale analysis was qualitatively compared with radar reflectivity, Tbb image to study its capability of reproduction of mesoscale systems. The results proved that the LAPS mesoscale analysis data had a strong reproduction capability on severe mesoscale convective system.

Meanwhile, with the data of radiosonde stations as references, the mean deviation, root mean square error and their distribution with heights of LAPS mesoscale analysis field were calculated and analyzed. The analysis results showed that all analysis errors of geopotential height, temperature, relative humidity, wind direction and wind velocity of LAPS mesoscale analysis field were within the range of observational errors. In particular, the average analysis error of geopotential height between the 700 and 925hPa was 45m for LAPS mesoscale analysis field; that of temperature was 1.0-1.1°C, that of relative humidity did not exceed 20%; that of wind velocity was 1.5 -2.0m/s, and that of wind direction was 20-25°.

The above results clearly indicated that the LAPS mesoscale analysis field possessed very high value in use and sufficient reliability, and it was quite applicable to mesoscale analysis.

VERIFICATION OF AERODROME THUNDERSTORM FORECASTS AND WARNINGS FOR THE HONG KONG INTERNATIONAL AIRPORT

P. Cheung, S.Y. Tang, C. C. Lam

Hong Kong Observatory, Hong Kong, China, e-mail: picheung@hko.gov.hk

Thunderstorms have great impact on the operation of both aircraft and aerodromes. Heavy thunderstorms cause significant delays and sometimes even safety issues to busy air hubs. There is, however, no provision in ICAO Annex 3 on how thunderstorm forecasts should be verified. In developing verification suite for assessing the performance of thunderstorm forecasts and warnings for the Hong Kong International Airport (HKIA), the Hong Kong Observatory (HKO) found episode-based verification coupled with duration-based verification simple yet comprehensive. In duration-based verification, forecast and warnings are verified by fixed time intervals (hour-by-hour for aerodrome forecasts and minute-by-minute for warnings) against the occurrence of thunderstorms. In episode-based verification, both forecasts and observations are segregated into “event episodes” and conventional performance indicators (POD, FAR, CSI, etc.) are derived based on automatic time matching of these episodes. The latter method has the advantage that other performance aspects such as forecast/warning lead time can be revealed. On the other hand, duration-based approach, though more stringent, reflects the gaps between forecasting skills and the desirable “perfect” forecast (exact timing and location). This paper presents the details of the verification for TAF and Aerodrome Warning of thunderstorms and how to interpret the verification results in order to identify areas of improvements.

AN EVALUATION OF MEASURES FOR EXTREME EVENTS AS PART OF THE SRNWP-V PROJECT

Matthew Trueman and Marion Mittermaier

Met Office, Exeter, UK, e-mail: matthew.trueman@metoffice.gov.uk

The main objective of the SRNWP-V (Short Range Numerical Weather Prediction-Verification) programme is to make an inter-comparison of the performance of the main operational limited-area SRNWP models. Currently there are 4 reference models in the inter-comparison: HIRLAM reference (FMI), Aladin-France, COSMO-DWD, the 12 km UM, and the ECMWF high resolution operational global model. Forecasts initialised at 00Z are verified as a function of forecast range (not availability).

The project is now in its extended phase. The first phase saw the verification of the following parameters: pmsl, 2m temperature, 2m relative humidity, 10m winds and 6h total precipitation against station observations, over the common area (in this case Aladin-France). The project has seen the establishment of a common framework for the exchange and verification of forecasts from representative reference versions of the main limited area operational models. Results have been produced since January 2009.

Part of this extended phase is to provide an evaluation of the usefulness of a number of scores developed for extreme events. Strong wind and heavy rainfall have been identified as the high-impact events of most interest. These two parameters will be evaluated by considering a number of thresholds to assess robustness. Given the comparatively long time series some inter-seasonal and inter-annual trends can be considered. In addition, we will also be investigating how climatologically derived thresholds could be used to define extreme events for stratified areas.

VERIFICATION OF CLOUD FORECAST FROM GLOBAL AND REGIONAL MODELS WITH SATELLITE DATA AT NCEP

Binbin Zhou[1,2], *Yuejian Zhu*[2] *and Geoff DiMego*[2]

[1]*I.M. System Group, Maryland, USA, e-mail:* Binbin.Zhou@noaa.gov
[2]*Environmental Modeling Center of NCEP/NOAA, Maryland, USA*

Cloud has significant impact on forecast performance of other fields like temperature and precipitation. However, verification of cloud forecast from NCEP's operational models has long been behind those of other regular fields due to a lack of appropriate validation cloud data and a practical verification tool. Verification of cloud with surface station data at NCEP has been conducted over CONUS. But for global domain, similar stationed cloud dataset is still not available for the operational weather center such as NCEP. Recently, with great efforts by NOAA/NESDIS and AFWA, satellite cloud dataset has been successfully retrieved and operationally transferred to NCEP for data assimilation and verification. Based on the NCEP's Grid-to-Grid Forecast Verification System (g2g), the cloud forecast from global (GFS) and regional model (NAM) now can be evaluated in real time since both satellite cloud data and forecast cloud data are in gridded data (in GRIB1 or GRIB2 format). In this paper, a brief introduction to the g2g and its application in verification of total cloud amount forecast from GSF and NAM will be presented. The results show that the forecast skills of cloud fraction from both GFS and NAM are still limited by comparing those regular fields.

NEAR REAL-TIME VERIFICATION OF CLOUD AND CONVECTION FORECASTS USING CLOUDSAT-CALIPSO GLOBAL RADAR-LIDAR OBSERVATIONS

Alain Protat, S. Young, L. Rikus, P. May, and M. Whimpey

CAWCR, Melbourne, Australia, email: a.protat@bom.gov.au

The way clouds are represented in numerical weather prediction (NWP) models significantly affects the quality of weather forecasts. The representation of deep convective systems and associated rainfall is also still a major challenge for NWP models, as there is growing evidence that NWP models produce way too many events but characterized by too little rainfall. The A-Train satellite mission offers new and unique opportunities to evaluate NWP models at global and regional scales. As part of the A-Train, CloudSat and CALIPSO provide a vertically-resolved description of the geometrical and microphysical properties of clouds and convection, which is crucial for model evaluation. In the present paper we describe a platform for the evaluation of clouds and convection generated by the regional and limited-area versions of the operational Australian NWP model (ACCESS) using near real-time observations of the CloudSat and CALIPSO mission from which radar-lidar cloud and convection masks are directly produced. Different skill scores calculated using one year of ACCESS forecasts will also be shown and discussed.

VERIFICATION OF HIGH RESOLUTION CLOUD COVER FORECASTS USING SATELLITE OBSERVATIONS

Ric Crocker and Jan Maksymczuk

Met Office, Exeter, United Kingdom, email: ric.crocker@metoffice.gov.uk

Surface based observations of cloud have several shortcomings when used for model verification. These shortcomings include areas of limited spatial coverage, for example over the sea, and limitations on the vertical column sampled by Cloud Base Recorders, meaning that high cloud is not always observed by automated observing systems. A project at the UK Met Office aims to use the Meteosat second generation (MSG) derived cloud mask as a verification source. The satellite cloud mask has full spatial coverage over the UK model domain and provides a top-down view ensuring that high clouds are detected. Results from this assessment can then be used to complement verification results obtained using surface observations, giving a fuller picture of model performance when forecasting cloud. A comparison of the satellite cloud mask at single sites against manual and automated synoptic observations is presented to demonstrate the differences between the observation types. Subsequently cloud cover forecasts from three configurations of the Met Office Unified Model (UM) have been assessed against the satellite cloud mask over a 5 month period. The focus at this stage is on how well the models forecast clear sky areas (cloud holes), the accuracy of which can influence other model parameters such as temperature and precipitation. The models have been assessed using several fuzzy verification techniques as well as the SAL feature-based method. Initial results have identified behavioural differences between the three models; particularly in the way that cloud fraction is distributed.

BENCHMARKING AND ASSESSMENT OF HOMOGENISATION ALGORITHMS FOR THE INTERNATIONAL SURFACE TEMPERATURE INITIATIVE (ISTI)

Ian Jolliffe and the ISTI Benchmarking and Assessment Working Group

University of Exeter, UK, e-mail: i.t.jolliffe@exeter.ac.uk

The International Surface Temperature Initiative (ISTI – www.surfacetemperatures.org) is an ambitious multi-disciplinary project, which aims to facilitate transparent creation of multiple long, high resolution, traceable (to source and known standards) data-products that are robust to varying non-climatic influences. As part of this a Benchmarking and Assessment working group (http://www.surfacetemperatures.org/benchmarking-and-assessment-working-group) will attempt to create a programme for objective assessment of data-products, specifically their ability to cope with non-climatic inhomogeneity.

Three task teams have been set up to guide key components: creation of synthetic realistic homogenous analog-known-worlds of data that mirror the ISTI global databank of surface temperature station records in format and station properties (location, climate mean state etc.); design of realistic error structures (introduction of inhomogeneities) to corrupt the analog-known-worlds creating analog-error-worlds; and design of analytical methods for assessing the ability of algorithms to homogenise the analog-error-worlds. Importantly, the analog-known-worlds and the error structures added must be physically realistic in character.

Although not *forecast* verification as such, the remit of the third team (Team Validation), namely comparison of homogenised data to analog-known-worlds, is equivalent. Assessment is needed of whether both the locations and nature of change-points in the analog-error-worlds are correctly identified and adjusted for. Verification will be required for single locations and for spatial areas, so that a wide range of verification methods will be needed. The exact nature of these methods will depend on the eventual form of the analog-error-worlds and their known-world counterparts, and there may a need for development of new verification methodology.

The intention is that there be three-year cycles, beginning with release of the analog-error-worlds, which data-product creators will attempt to homogenise, resulting in products that are as close as possible in some sense to the corresponding analog-known-worlds. The latter will be unknown to the data-product creators. The cycle will end with a workshop bringing together benchmark group and data-product creators to assess the properties of the homogenisation algorithms and improve the benchmarks for the next cycle.

CONCEPTS FOR A PATTERN-ORIENTED ANALYSIS ENSEMBLE BASED ON OBSERVATIONAL UNCERTAINTIES

Manfred Dorninger and *Theresa Gorgas*

University of Vienna, Department of Meteorology and Geophysics, Vienna, Austria,
e-mail: Manfred.Dorninger@univie.ac.at

When choosing the reference data representing the 'truth' for a specific verification study we must be aware that the results of several verifications statistics do not only depend on the forecast error, i.e. the information which is the aim of the verification study, but also on errors of the verification reference data itself.

To account for this effect the idea of an analysis ensemble of deterministic, model-independent analyses is proposed, based on random perturbations for irregularly distributed observations. The purpose of implementing an analysis ensemble is to define uncertainties in analysis fields due to their observational background and errors. As one possible application the uncertainty information could be used to define confidence intervals for verification measures depending on the reference data. The analysis system VERA (Vienna Enhanced Resolution Analysis) and a high resolution Central European observation network (Joint D-PHASE-COPS data set) are used as testbed for the development of the methodology. Several approaches for defining reasonable weights for the perturbation fields are investigated and compared. Basic weights are determined by a sophisticated data quality control scheme producing error estimates for observations. They can be combined with additional information attempting to include the spatial structure of the observed fields in the ensemble more explicitly. The information is provided by either a principal component analysis of a time series of analysis fields or a 2D-discrete wavelet transform.

Strengths and weaknesses of the different adjustments for ensemble analysis perturbations are discussed. It is shown that perturbations provided by the wavelet based approach lead to useful results for several meteorological parameters. They reflect the spatial patterns in the perturbed fields and provide balanced distributions around the un-perturbed reference analysis, as is shown by basic statistics. The advantages of the wavelet transform for ensemble objectives are twofold: It is fast and the results can be interpreted physically.

RADAR RAINGAUGE VERIFICATION

Sandy Dance and Alan Seed

CAWCR, Melbourne, Australia, e-mail: s.dance@bom.gov.au

In this paper we analyse the relationship between radar and raingauge estimations of rainfall in order to elucidate the error structure of radar rainfall accumulations. We describe how we generate simulated gauge rainfall accumulations from radar data rainfall accumulations that accumulate over shorter, offset periods. We then show exemplar time series of rainfall accumulations from radar and gauge, indicating that we are successfully simulating gauge data.

We analyse the statistical nature of radar - gauge errors, and conclude that the correct error measure is log(radar/gauge). We then try to find correlates of the error variance in terms of terrain topography, radar range, and season.

The Centre for Australian Weather and Climate Research is a partnership between CSIRO and the Bureau of Meteorology.