Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Annalen der Meteorologie 51

7th International Verification Methods Workshop

03. - 11. May 2017
in Berlin, Germany

Programme and Abstracts

Offenbach am Main 2017
Selbstverlag des Deutschen Wetterdienstes

Annalen der Meteorologie **51**

7th International Verification Methods Workshop

03. - 11. May 2017
in Berlin, Germany

Programme and Abstracts

Hans-Ertel-Zentrum für Wetterforschung

Freie Universität Berlin

JWGFVR

World Meteorological Organisation (WMO)

World Weather Research Programme (WWRP)

World Climate Research Programme (WCRP)

Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Offenbach am Main 2017
Selbstverlag des Deutschen Wetterdienstes

Conference papers online:
www.dwd.de/EN/ourservices/pbfb_verlag_annalen/annalen.html

Workshop homepage:
www.7thverificationworkshop.de/index.html

ISSN der Online-Ausgabe: 2194-5934
ISSN der Druckausgabe: 0072-4122

ISBN der Online-Ausgabe: 978-3-88148-495-4
ISBN der Druck-Ausgabe: 978-3-88148-496-1

Herausgeber und Verlag:

Deutscher Wetterdienst
Fachinformationsstelle und
Deutsche Meteorologische Bibliothek
Frankfurter Straße 135
63067 Offenbach am Main
www.dwd.de/bibliothek
bibliothek@dwd.de

Redaktion:

Martin Göber
Deutscher Wetterdienst
Forschung und Entwicklung
Frankfurter Str. 135
63067 Offenbach am Main
Martin.Goeber@dwd.de

Druck und Bindung:

Druckerei des BMVI
Robert-Schumann-Platz 1
53175 Bonn

Foreword

“Forecast verification methods across time and space scales“

On behalf of the organising committee of the 7th International Verification Methods Workshop we would like to welcome you in Berlin, Capital of Germany, a place of an eventful history, of great cultural treasures, scientific discoveries and technological inventions, a city with a vibrant life and a diverse and tolerant population.

After Boulder, Montreal, Reading, Helsinki, Melbourne and New Delhi the Seventh in the very successful series of verification methods workshops has returned to Europe, to bring together almost 40 students and about 80 presenters from 35 countries from all continents from national and private weather services, from universities and research institutions and intergovernmental organisations to teach and learn and discuss and enjoy the growing multitude of methods of verification of weather forecasts and warnings, climate predictions, and their applications.

The workshop starts with a tutorial on forecast verification methods (May 3 - 6) with lectures and hands-on laboratory sessions using the R statistical language. It will be hosted by the Max-Planck-Institute for Human Development.

The workshop continues with the science conference (May 8 - 11) including keynote talks along with contributed presentations and posters. This will take place in the „VKU Forum“ in Berlin-Mitte.

The workshop has been organised by the WMO Joint Working Group on forecast Verification Research, jointly with the Free University of Berlin, the Max-Planck-Institute for Human Development, the Hans-Ertel-Centre for Weather Research (HErZ) and the German Weather Service DWD.

In this booklet you will find the schedule of the science conference, the abstracts of the talks and posters, a description of the venue, how to get lunch at the university canteen (the „Mensa“) and how to find the way to the start of the conference dinner - a boat trip through the city of Berlin!

We forecast days of great talks, posters, discussions, spirits, food and weather - please verify!

Martin Göber

Content

1. Venue **6**
1.1 Tutorial Venue 6
1.2 Conference Venue 7

2. Public Transport **9**

3. Lunch **10**

4. Ice Breaker/Conference Dinner **11**

5. Timetable **12**

6. Oral Presentations **19**

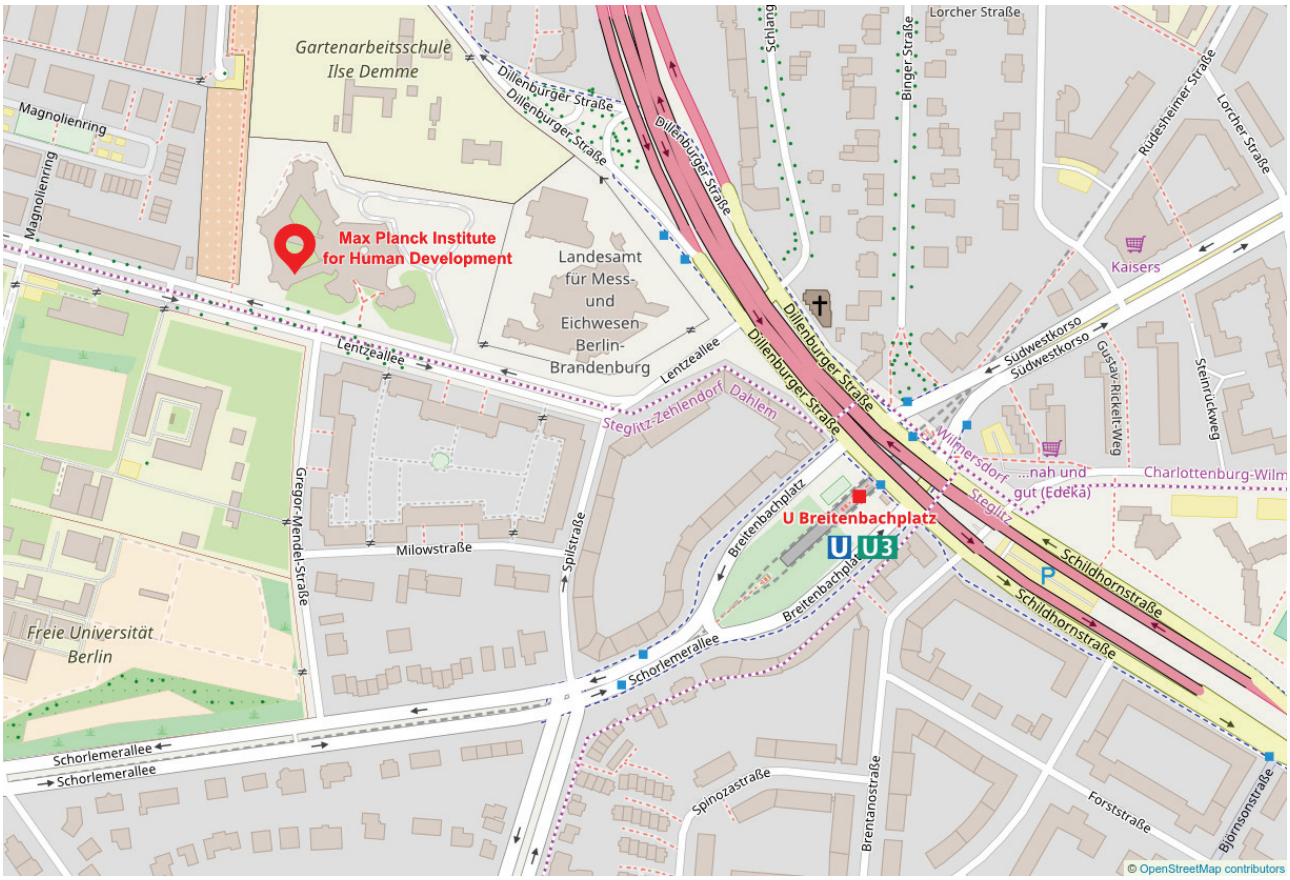
7. Poster Presentations **45**

1. Venue

1.1 Tutorial Venue

The tutorial will take place at the

Max-Planck-Institute for Human Development (MPIB)
Lentzeallee 94
D-14195 Berlin



The MPIB is a 5-minute walk from the underground station Breitenbachplatz.

Public Transport

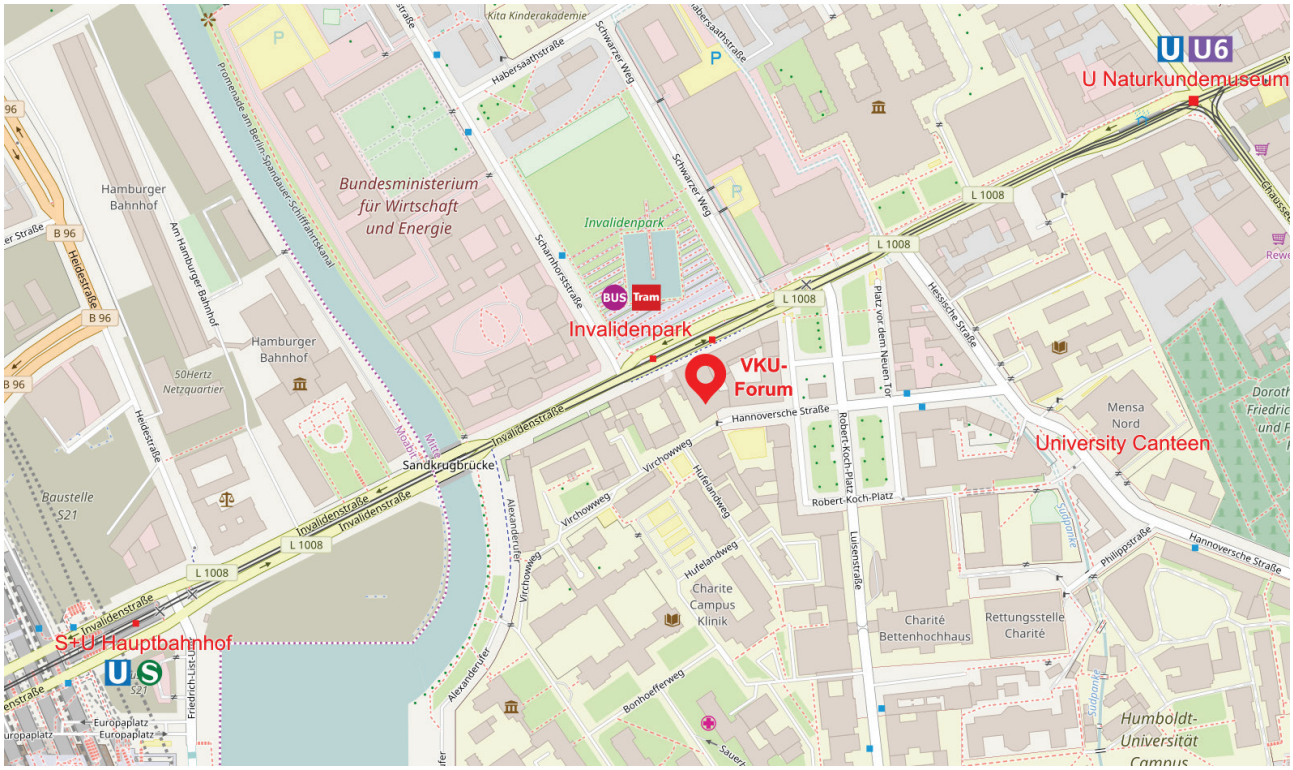
From station Zoologischer Garten:

Take the underground (U-Bahn) line U9 going to Rathaus Steglitz. At the subway station Spichernstraße get off and change to the subway line U3 going to Krumme Lanke and get off at Breitenbachplatz station.
The MPIB is a 5-minute walk from Breitenbachplatz (see map).

1.2 Conference Venue

The conference will take place at the

VKU Forum
Invalidenstraße 91
D-10115 Berlin



Public Transport

<u>Arriving by plane</u>

From Tegel Airport:

Take the Bus TXL in direction S+U Berlin Hauptbahnhof and get of at the stop Invalidenpark which is just in front of VKU Forum (about 20 minutes).

From Schönefeld Airport:

Take either the regional train (Regionalbahn) RB7 in direction Dessau, RB14 in direction Nauen or RB19 in direction S+U Gesundbrunnen Bhf and get of at the stop S+U Berlin Hauptbahnhof. Than take any of the options listed below!

Arriving by train

From Berlin-Hauptbahnhof (main railway station):

Take one of the following services and get of at the stop Invalidenpark (about 2 minutes)

- Tram M5 in direction Zingster Straße
- Tram M8 in direction Ahrensfelde/Stadtgrenze
- Tram M10 in direction S+U Warschauer Straße
- Bus 123 in direction Robert-Koch-Platz
- Bus 142 in direction S Ostbahnhof
- Bus 147 in direction S Ostbahnhof
- Bus 245 in direction Robert-Koch-Platz

Otherwise, it is a 5-10 minutes walk from Berlin-Hauptbahnhof (see map). From the exit “Europaplatz” turn right and follow „Invalidenstraße“ for about 600 m.

From station Zoologischer Garten:

Take Bus 245 or the commuter train (S-Bahn) S5 in direction Hoppegarten or S75 in direction Wartenberg to Berlin-Hauptbahnhof, then take one of the services above.

2. Public Transport

For all information about public transport in Berlin by commuter train (S-Bahn), underground (U-Bahn), Tram and Bus with journey planner, citymap, routemap, ticket and traffic information see http://www.vbb.de/en/index.html. **A (copyrighted) map of the city trains and underground can be found here, which you might find useful to download: http://images.vbb.de/assets/downloads/file/1193159.pdf**. The city of Berlin and the surrounding suburbs have been organized into three tariff zones A, B and C. You can buy tickets for the combined zones AB, BC or ABC, as needed. **Normally you need Berlin AB**.
Adults pay regular fare. The different forms of tickets are: single – day – weekly.

Ticket	Berlin AB	Prices
	Single ticket	2,80 €
	Single ticket 4-trip-ticket	9,00 €
	Day ticket	7,00 €
	7-Day-ticket	30,00 €

Single Tickets must be validated immediately on starting travel. These tickets are not transferable after the journey has been started. They are valid for a maximum of two hours. Within the two-hour validity of the Single Ticket you may change trains/buses etc. or interrupt your journey as often as you wish. Return trips (back to the starting point) and round trips (journeys that take a different route from the outbound journey) are not permitted. Single tickets are available also as **4-trip-ticket**.

Day tickets must be validated on starting travel. They are not transferable after the journey has been started. With a **Day ticket** you can travel as often as you want on the day printed on the ticket or from validation of the ticket on starting your journey until 3.00 a.m. on the following day.

7-Day-Tickets must be validated before starting your first journey. The 7-Day-Ticket is valid for seven consecutive calendar days. Validity begins on the first day of the validity period printed on the ticket or from the time of validation and ends on the seventh calendar day, at midnight. For example, from validation at 9.30 a.m. on a Tuesday until midnight on the following Monday. The Ticket is transferable and can also be used from other persons.

Tickets can be purchased conveniently at one of the many sales points or at the ticket machines that can be found at all stations, transfer points, and inside the trams or directly from the bus driver. **Please do not forget to validate your ticket before starting your journey!** Ticket validation machines are located on all platforms at the S-Bahn and U-Bahn stations and in case of trams and buses inside the vehicles.
It is not necessary to validate tickets that are purchased directly from the driver in buses. Please board buses through the forward door; pre-purchased tickets must be displayed to the driver upon entry.

Arriving at Berlin's airports

Airport Tegel

Bus connections to Berlin city centre	direction
Express bus TXL	Alexanderplatz (city center east)
Express bus X9	S + U Zoologischer Garten (city center west)
Bus 109	S + U Zoologischer Garten
Bus 128	U Osloer Straße

Airport Schönefeld

The Airport Express – just 28 minutes from Central Station to Schönefeld Airport. Specially marked regional train lines RE7 and RB14 run daily every 30 minutes (5am - 11pm) between several city stations and the airport. Local public transport tickets are accepted on these trains. The journey is within fare zone ABC.

S-Bahn:	Lines S45 and S9
Buses:	SXF1, X7, 162, 163, 171, N60, N70

3. Lunch

A lunch opportunity at the university canteen “Mensa Nord”, 7 minutes walk from the conference venue (see map conference venue)

What to eat?
The daily menu you can check before at:
https://www.studentenwerk-berlin.de/en/mensen/speiseplan/hu_nord/index.html

How to pay?
You can make cashless payments with the MensaCard. The MensaCard is a rechargeable smart card on which only the amount and card number are stored. Personal data such as names, frequency of use or personal preferences for certain foods are not stored at any point.
The MensaCard for use during the conference days **will be handed out at the conference registration on Monday.**
PLEASE have 10 EURO available, it is pre-loaded with an amount of 8,45 EURO (1,55 EURO deposit, paid back at return).

Please give the MensaCard back at the registration desk when leaving the conference.

Place the card flat on the reading area of the card reader at the checkout after choosing your meal.
The balance is shown at the top of the display. The amount to be deducted is shown at the bottom of the display. After deduction of the debit amount shown at the bottom, the new card balance is shown at the top of the display.

How are MensaCards recharged?
Cards are recharged at the recharging terminals set up for this purpose. They can be recharged by inserting bank notes with a value of 5, 10, 20 or 50 euro. The amount inserted is shown on the display. After inserting the money, hold the card directly in front of the card reader and wait until the new balance is shown. When it appears, the card has been successfully recharged. At facilities without a recharging terminal, you can recharge your MensaCard at the checkouts.

How can I check the balance on the card?
Hold the card in front of the card reader of a recharging terminal. The balance is shown on the display. Before choosing your meal, please check the balance on your MensaCard.

How can I return the card to receive the deposit?
The cards are simply returned at the last checkout at the respective checkout, and the credit balance and deposit are paid out.

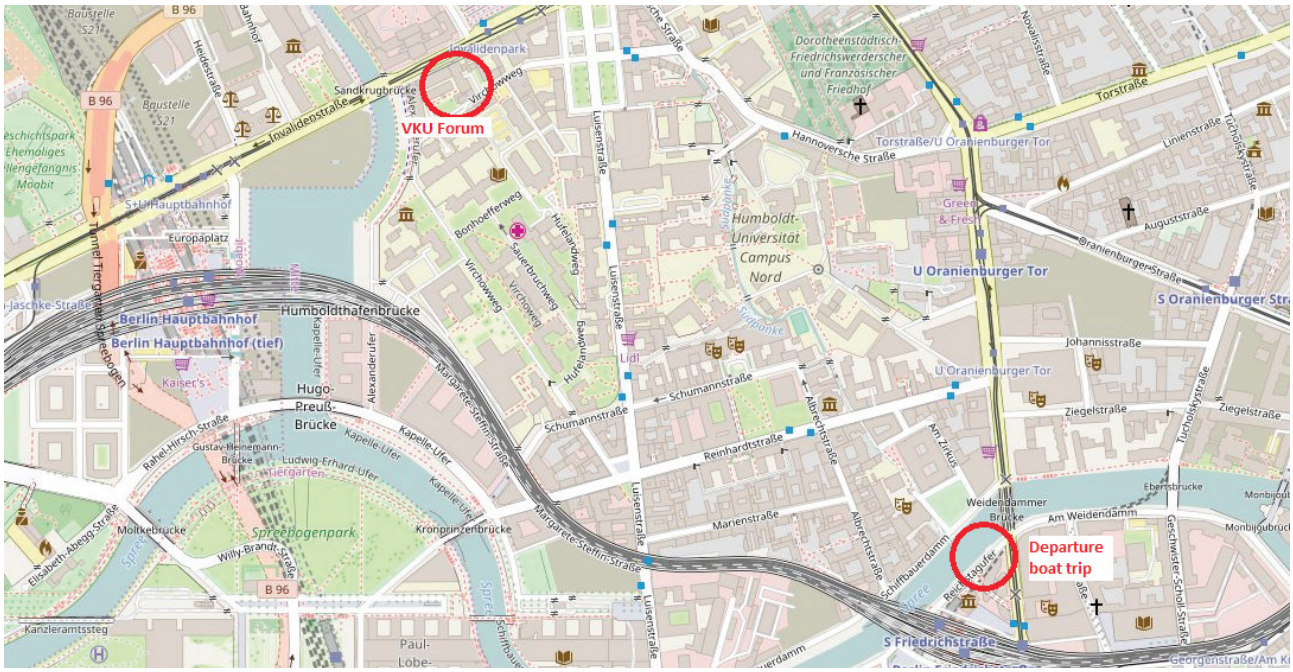
4. Ice Breaker/Conference Dinner

There will be two parties during the conference: the ice breaker party at the end of the program on Monday at the venue and the conference dinner in the evening of Wednesday (see Boat trip).

Boat-trip
The conference dinner in the evening of Wednesday, May 10, will be combined with a boat trip underneath the bridges of the Spree passing by the sights of Berlin.

Departure landing stage:	Friedrichstraße (address: Reichstagufer 18, in front of the Tränenpalast, north of the train station Friedrichstraße)
Departure:	7 pm (**boarding starts 25 min. prior to departure**)
Return:	11 pm

Please don't forget the ticket for the dinner, which will be handed out to you at the registration.



Public transport

From Invalidenpark (VKU)
BUS 147 in direction S Ostbahnhof → S/U Friedrichstraße

From U Naturkundemuseum
Take one of the following services and get of at the stop S/U Friedrichstraße

- U U6 in direction U Alt-Mariendorf
- Tram 12 in direction Mitte, Am Kupfergraben

Otherwise, it is a 25-30 minutes walk from VKU (1,5 km). From the exit VKU turn right and follow "Invalidenstraße" for about 200 m. Turn right into "Hannoversche Straße", after 500 m turn right into "Friedrichstraße", after 700 m on the other side of the river "Spree", right from the "Weidendammer Brücke" you will see the landing stage ("Bruno Winkler").

5. Timetable

Sessions

HIW	Verification of high impact weather and extremes
META	Meta-verification, properties of verification methods
OBS UNC	Observation uncertainty
USR VAL	User-oriented verification and estimation of forecast value
PROB ENS	Methods for verification of probability and ensemble forecasts
SPATIAL	Spatial verification methods and verification approaches
OPS SOFT	Software and measures used to evaluate operational predictions, including verification methods used to justify upgrades to operational models.
S2S	Sub-seasonal to seasonal
CLIM	Climate
MISC	Verification studies

Time	Session	Title	First Name	Surname	Organization	Country
Monday						
10:00-11:00	Registration Poster hanging					
11:00-11:20	Welcome remarks					
11:20-12:00	HIW key-note	Verification for High Impact Weather	Beth	Ebert	Bureau of Meteorology	Australia
12:00-12:20	HIW	Verification of extremes based on extreme value theory	Petra	Friederichs	Meteorological Institute, University of Bonn	Germany
12:20-12:40	HIW	Verification of extremes using proper scoring rules and extreme value theory	Maxime	Taillardat	CNRM/Météo-France	France
12:40-14:00	**lunch**					
14:00-14:20	HIW	Exploring the forecasting and warning value chain in HIWeather	Brian	Golding	Met Office	UK
14:20-14:40	HIW	Development of verification methodology for extreme weather forecasts	Hong	Guan	SRG at EMC/ NCEP/NWS/ NOAA	United States

Time	Session	Title	First Name	Surname	Organization	Country
14:40-15:00	HIW	Forecaster's Dilemma: Extreme Events and Forecast Evaluation	Sebastian	Lerch	Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology	Germany
15:00-15:40	USR key note	Economic Assessment of Hydro-Met Services and Products: A Value Chain Approach	Jeff	Lazo	National Center of Atmospheric Research	USA
15:40-16:00	**coffee break**					
16:00-16:20	USR	A user-oriented verification methodology for wind forecast.	Maria Stefania	Tesini	Arpae-SIMC	Italy
16:20-16:40	USR	User-oriented evaluation of fire spread predictions	Beth	Ebert	Bureau of Meteorology	Australia
16:40-17:00	USR	Develop a user oriented forecast verification metric	William (Xiang-dong)	Wang	Australian Bureau of Meteorology	Australia
17:00-17:20	USR	The final approach? Verification of Terminal Aerodrome Forecasts	Michael	Sharpe	Met Office	UK
17:20-17:40	USR	Displaying the Verification Results of Terminal Aerodrome Forecasts for Thunderstorms and Visibility	Jadran	Jurkovi	Croatia Control Ltd	Croatia
17:40-18:00	Discussion HIW & USR	Facilitators: Beth Ebert, Jeff Lazo, Brian Golding				
18:00-22:00	**Ice breaker**					

Time	Session	Title	First Name	Surname	Organization	Country
Tuesday						
8:20-9:00	META key note	Forecast verification using scoring rules	Chris	Ferro	Exeter University	UK
9:00-9:20	META	Proper and equitable scores: a resolved dilemma	Zied	Ben Bouallegue	ECMWF	UK
9:20-9:40	META	The generalized discrimination score: connections, corrections and potential applications	Roger	Harbord	Met Office	UK
9:40-10:00	META	Decomposition and Attribution of Forecast Errors	Fanglin	Yang	Environmental Modeling Center, National Centers for Environmental Prediction	USA
10:00-10:40	**coffee break**					
10:40-11:00	META	Murphy diagrams	Alexander	Jordan	Heidelberg Institute for Theoretical Studies	Germany
11:00-11:20	META	Verification of Multi-Valued Forecasts	Tim	Bullock	Meteorological Service of Canada	Canada
11:20-11:40	ENS	Ensemble verification: Old scores, new perspectives	Sabrina	Wahl	Hans-Ertel-Centre for Weather Research / University of Bonn	Germany
11:40-12:00	ENS	Increasing the usability of analysis rank histograms	Jan	Keller	Hans-Ertel-Centre for Weather Research / Deutscher Wetterdienst	Germany
12:00-12:20	ENS	Another look at spread and skill	Åke	Johansson	Swedish Meteorological and Hydrological Institute (SMHI)	Sweden
12:20-12:40	Discussion META ENS	Facilitators: Chris Ferro, Barb Brown, Chiara Marsigli				
12:40-14:00	**lunch**					

Time	Session	Title	First Name	Surname	Organization	Country
14:00-14:20	ENS-spatial	The development of probabilistic forecast verification across the scales	Yuejian	Zhu	Environmental Modeling Center, NCEP/ NWS/NOAA	U. S. A
14:20-14:40	ENS-spatial	Generating and verifying probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations	Craig	Schwartz	National Center of Atmospheric Research	USA
14:40-15:00	ENS-spatial	Benefit of an ensemble over a deterministic forecast at km-scale	Marion	Mittermaier	Met Office	UK
15:00-15:20	OPS-inference	The difficulty of verifying small improvements in forecast quality	Alan	Geer	ECMWF	UK
15:20-15:40	OPS-inference	Methods for Identifying Small Improvements in Forecasts	Elizabeth	Weatherhead	University of Colorado	USA
15:40-16:00	Discussion ENS inference	Facilitators: Chris Ferro, Marion Mittermaier, Chiara Marsigli				
16:00-16:20	**coffee break**					
16:00-18:00	poster session			all		

Time	Session	Title	First Name	Surname	Organization	Country
Wednesday						
8:30-8:40	intro verif challenge	The verification challenge	Beth	Ebert	Bureau of Meteorology	Australia
8:40-9:20	Key note: verification challenge winner	Verification of user-relevant contours with the Integrated Ice Edge Error and the Spatial Probability Score	Helge	Goessing	Alfred Wegener Institute	Germany
9:20-9:40	SPATIAL	Verification of sea-ice prediction by using distance measures	Barbara	Casati	Meteorological Research Division, Environment and Climate Change Canada	Canada
9:40-10:00	SPATIAL	Comparing distance methods for spatial verification	Eric	Gilleland	National Center for Atmospheric Research	U.S.A.
10:00-10:40	**coffee break**					
10:40-11:00	SPATIAL	Estimating the displacement in precipitation forecasts using the Fractions Skill Score	Gregor	Skok	Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia	Slovenia
11:00-11:20	SPATIAL	A multi-model, multi-analysis study to asses the capability of CRA analysis for QPF spatial verification in the MesoVICT framework	Stefano	Mariani	ISPRA - Institute for Environmental Protection and Research	Italy
11:20-11:40	SPATIAL	Evaluation of Extra-Tropical Cyclones Using a Feature-Relative Method	Tara	Jensen	National Center for Atmospheric Research/ Research Applications Lab	USA
11:40-12:00	SPATIAL	Methods for Evaluation of Cloud Predictions	Barbara	Brown	NCAR	United States
12:00-12:20	SPATIAL	Grid-based spatial verifications versus point-based verifications with the application to precipitation and clouds	Johannes	Jenkner	UBIMET GmbH	Austria

Time	Session	Title	First Name	Surname	Organization	Country
12:20-12:40	Discussion SPATIAL	Facilitators: Manfred Dorninger, Eric Gilleland, Barb Brown				
12:40-14:00	**lunch**					
14:00-14:20	SOFT-ops-spatial	A Unified Approach to Verification	Tara	Jensen	National Center for Atmospheric Research/ Research Applications Lab	USA
14:20-14:40	OBS UNC	Evaluating representativeness errors in verification against Arctic surface observations	Thomas	Haiden	ECMWF	United Kingdom
14:40-15:00	OBS UNC	Spatial verification of ensemble forecasts with SAL and the role of observation uncertainty	Sabine	Radanovics	LSCE	France
15:00-15:20	OBS UNC	Considering observation uncertainty for verification of forecast ensembles	Manfred	Dorninger	University of Vienna	Austria
15:20-15:40	OBS UNC	Impact of gridded truth resolution on spatial verification scores“	Marion	Mittermaier	Met Office	UK
15:40-16:00	Discussion OBS UNC	Facilitators: Thomas Haiden, Manfred Dorninger, Marion Mittermaier				
16:00-16:20	**coffee break**					
16:20-18:00	tutorial project presentation					
19:00-23:00	**Conference Dinner: Boat trip**			**all**		

Time	Session	Title	First Name	Surname	Organization	Country
Thursday						
8:20-9:00	CLIM key note	Drift in Decadal Prediction: a Particular Challenge for Verification	Henning	Rust	Freie Universität Berlin	Germany
9:00-9:20	CLIM	Freva - Freie Univ Evaluation System Framework for Scientific Infrastructures in Earth System Modeling	Christopher	Kadow	Freie Universität Berlin	Germany
9:20-9:40	CLIM	Joint uncertainty assessment of models and observations in verification of climate predictions.	Omar	Bellprat	Earth Sciences, Barcelona Supercomputing Centre (BSC)	Spain
9:40-10:00	CLIM	Useful evaluations of variability and change in simulated climate for impact studies	Caspar	Ammann	National Center for Atmospheric Research	United States
10:00-10:40	**coffee break**					
10:40-11:20	S2S key note	Sub-seasonal to seasonal forecast Verification	Frederic	Vitart	ECMWF	United Kingdom
11:20-11:40	S2S	Evaluation and Quality Control for the Copernicus Seasonal Forecast Systems	Jonas	Bhend	Federal Office of Meteorology and Climatology MeteoSwiss	Switzerland
11:40-12:00	S2S	Evaluating Subseasonal hindcast using an objective regionalization approach	Gloria	Recalde	Department of Earth and Planetary Science, Johns Hopkins University	United States
12:00-12:20	Discussion CLIM S2S	Facilitators: Caio Coelho, Frederic Vitart, Henning Rust				
12:20-13:00	**Workshop wrap-up discussion**					
13:00-onwards	**lunch and networking**					

6. Oral Presentations

HIW-1 KEY NOTE: Verification for High Impact Weather

AUTHORS: Ebert, Beth.

Presenting author e-mail: beth.ebert.@bom.gov.au; Affiliation (all authors): Bureau of Meteorology, Melbourne, Australia

ABSTRACT: Timely and accurate forecasts for high impact weather can prompt people to take protective action. Understanding the forecast quality is critical to using it effectively, so high impact weather forecasts must be verified in ways that are meaningful to users in the public, emergency management, government, industry, and other sectors. In 2016 WMO kicked off a 10-year High Impact Weather (HIWeather) project to improve forecasts on timescales of minutes to weeks and enhance their utility in social, economic and environmental applications. HIWeather activities are focused around five weather hazards: urban flood, wildfire, localised extreme wind, disruptive winter weather, and urban heat waves and air pollution. Forecasts and warnings for these types of extreme events usually depend on numerical weather prediction at lead times beyond a few hours and nowcasts at shorter time ranges, with ensemble and probabilistic approaches often used to account for forecast uncertainty. While this reflects scientific best practice it immediately throws up a challenge for verifying individual events where people want to know whether or not the (probabilistic) forecast was correct. Good observational data for extreme events may be hard to obtain and often contain significant uncertainties in timing, location, and magnitude. New verification approaches are being developed to explicitly account for uncertainty in the observations. High impact weather verification has tended to use the same statistical approaches that are used for everyday weather forecasts. Categorical extreme dependency scores show better behaviour for rare extreme events than traditional threat and skill scores and are coming into common use, especially for assessing competing forecast systems. Spatial verification approaches are also widely used for high resolution forecasts to give credit to close forecasts. To make verification more meaningful for users outside the weather community, new scores have been proposed that are more relevant to the way people make their decisions. As we move toward predicting hazard impacts such as storm damage, poor health outcomes, etc., an important topic of research will be how best to collect and use new types of observations, including from social media, in verification. This talk will highlight challenges that come with verifying forecasts for high impact weather. It will describe some of the activities being undertaken in the HIWeather project, and show examples of new verification approaches that are being applied to evaluate forecasts of weather hazards.

HIW-2: Verification of extremes based on extreme value theory

AUTHORS: Friederichs, Petra; Wahl, Sabrina.

Presenting author e-mail: pfried@uni-bonn.de; Affiliation (all authors): Meteorological Institute, University of Bonn, Bonn, Germany

ABSTRACT: Extremes are per definition rare events, and thus standard verification approaches often provide only degenerated measures in the limit of extremes. This is partially due to different convergence rates of the respective measures. Extreme value theory (EVT) provides a limit law for extremes, and thus a methodology to mathematically describe the behavior of extremes in a quite universal way. Chris Ferro (2007) makes use of limit laws in EVT to provide estimates of entries in the contingency table in case of extremes. We follow this line and explore verification approaches using EVT for univariate, multivariate or spatial extremes.

HIW-3: Verification of extremes using proper scoring rules and extreme value theory

AUTHORS: Taillardat, Maxime; Fougères, Anne-Laure; Naveau, Philippe; Mestre, Olivier.

Presenting author e-mail: maxime.taillardat@meteo.fr; Affiliation (all authors): CNRM/Météo-France, Lyon, France

ABSTRACT: In meteorology, quantitative verification of extremes is quite difficult. Hedging or weighting strategies can be employed, especially with some versions of the Continuous Ranked Probability Score (CRPS), but it can lead to improper or degenerative scores. We propose here to assess verification of extremes focusing on CRPS's distribution. Studying the behaviour of the CRPS for extreme events and using extreme value theory, we make a link between extreme values of the CRPS and the CRPS for extreme events. We exhibit a new quantity for the verification of extreme events, based on the CRPS, which can be estimated by leveraging extreme value theory.

HIW-4: Exploring the forecasting and warning value chain in HIWeather

AUTHORS: Golding, Brian.

Presenting author e-mail: brian.golding@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: The WMO High Impact Weather (HIWeather) project has been established to promote research in areas of physical and social science that will improve forecasting and warning systems for weather-related hazards. The end result of any warning system is that a population of warning recipients makes a set of decisions that may reduce suffering and economic loss. It is increasingly important that forecasting and warning systems can demonstrate that investment will lead to increased value, in terms of reduced suffering and loss. In this talk, I will explore the nature of the forecasting and warning system; the relationship between skill and value; where and why value is gained and lost; and what sorts of measurements might contribute in an end-to-end assessment of the value chain. Examples will be given of candidate metrics at each stage of the value chain, emphasising the challenges in obtaining validation data, in achieving statistical significance and in relating the scores to each other.

HIW-5: Development of verification methodology for extreme weather forecasts

AUTHORS: Guan, Hong; Zhu, Yuejian.

Presenting author e-mail: hong.guan@noaa.gov; Affiliation (all authors): SRG at EMC/NCEP/NWS/NOAA, College Park, United States

ABSTRACT: In past years, the forecast skill has been improved significantly from many areas. Those are observation system, data assimilation, numerical modeling, and ensemble forecasting. Today, ensemble forecast routinely provides essential information on the probability of specific events. A predictive skill now extends in some case beyond 10 days, with an increasing capability to give early warning of severe weather events many days ahead. One of North American Ensemble Forecast System (NAEFS) applications is producing anomaly forecast (ANF) from bias corrected NCEP and CMC's global ensemble forecasts and analysis climatology since 2006. Later on, the extreme forecast index (EFI) has been introduced to predict extreme weather events those based on raw ensemble forecast and model climatology. Although the extreme weather forecast products, such as ANF and EFI, have been generated in real time daily operation, and the applications are widely used in many areas, the verification of the products has been a challenge. Without valuable verification tool, the performance of these products is still uncertain, especially for ensemble upgrade. In this work, we develop a verification method to evaluate the relative ability of the ANF and EFI in predicting extreme events. In this study, we quantify the relationship between ANF and EFI and find the equivalent threshold values for identifying extreme events from the two algorithms. Then we establish binary contingency table based on the selected threshold values for different applications to calculate statistic scores, such as hit rate, false alarm rate, frequency bias and equitable thresh score for each product. Finally, a previous developed visualization tool (performance diagram) is used to identify relatively better extreme weather forecast products. The verification methodology has been applied to examine extreme cold and extreme precipitation forecasts, the results will be presented as well.

HIW-6: Forecaster's Dilemma: Extreme Events and Forecast Evaluation

AUTHORS: Lerch, Sebastian; Thorarinsdottir, Thordis L.; Ravazzolo, Francesco; Gneiting, Tilmann.

Presenting author e-mail: Sebastian.Lerch@h-its.org; Affiliation (all authors): Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT: In public discussions of the quality of forecasts, attention typically focuses on the predictive performance in cases of extreme events. However, the restriction of conventional forecast evaluation methods to subsets of extreme observations has unexpected and undesired effects, and is bound to discredit skillful forecasts when the signal-to-noise ratio in the data generating process is low. Conditioning on outcomes is incompatible with the theoretical assumptions of established forecast evaluation methods, thereby confronting forecasters with what we refer to as the forecaster's dilemma. For probabilistic forecasts, proper weighted scoring rules have been proposed as decision-theoretically justifiable alternatives for forecast evaluation with an emphasis on extreme events. Using theoretical arguments, simulation experiments, and a case study on probabilistic wind speed forecasts, we illustrate and discuss the forecaster's dilemma along with potential remedies.

USR VAL-1 KEY NOTE: Economic Assessment of Hydro-Met Services and Products: A Value Chain Approach

AUTHORS: Lazo, Jeffrey K..

Presenting author e-mail: lazo@ucar.edu; Affiliation (all authors): Societal Impacts Program, National Center for Atmospheric Research, Boulder, United States

ABSTRACT: As many in the forecast verification community already know, weather related hazards such as typhoons, floods, heat waves, droughts, and tornadoes cause billions of dollars of damage and affect millions worldwide in both developed and developing countries. Between 2004 and 2013 an annual average of 127 meteorological, 32 climatological, and 192 hydrological disasters affected an average of 191 million people each year, and caused an average annual $122 Billion dollars of damages (Guha et al 2015). "Day-to-day" events (not considered "disasters") likely have an even larger aggregate impact on society and affect virtually everyone on the planet in some manner every year " if not every day! While not all (or perhaps even most) of the impacts can be avoided or mitigated, with appropriate information systems and processes there are undoubtedly significant societal benefits of geo-spatial information on weather, water, and climate. Developing a better understanding of the socio-economic value of hydro-met information is the focus of some recent efforts (WMO 2015). In this talk I'll present the concept of the "Weather Information Value Chain" as a tool for understanding the creation of value from hydromet information as well as explicating the difficulties of valuation and opportunities for value enhancement. Building on the information value chain concept, I offer that economics as a study of human behavior and decision making can help in understanding the weather related decision making process and thus enhance product development, communication approaches, and ultimately decision making to increase societal value. I'll discuss the relationship between forecast verification and economic benefit analysis and the developing concept of user-relevant verification. This topic will hopefully provide grounds for further discussion with the verification community as it is truly a concept in progress! Time permitting, I'll present several examples of valuation of hydromet information in the context of a "value chain" including a study on research to improve forecasting for utility scale solar power generation (Haupt et al. 2016). Finally I'll briefly note some factors related to explication through the value chain model including (1) alternative methods for characterization of the information value process (Lazo et al 2016), (2) the significant opportunity for the application of concepts from behavioral economics for better understanding responses to hydro-met information, (3) the context, concept, and measurement of vulnerability and resilience within which the Weather Information Value Chain is embedded, and (4) ethical frameworks that could be more explicitly articulated in the process of value assessment as well as in decision-making for improving weather information processes.

USR VAL-2: A user-oriented verification methodology for wind forecast.

AUTHORS: Tesini, Maria Stefania.

Presenting author e-mail: mstesini@arpae.it; Affiliation (all authors): Arpae-SIMC, Bologna, Italy

ABSTRACT: The 10-m wind is a weather parameter characterized by strong dependence on orographic and topographic details and high temporal variability. Therefore the verification of wind forecast requires a methodology taking into account these features. On top, any verification method should be tailored for the specific purpose defined by the user of that forecast, being the developer of the model, the forecaster in the operational room or the stakeholder for a practical application. In this work, the 10-m wind predicted by the COSMO model run at 7 and 2.8 km over Italy (COSMO-I7 and COSMO-I2 suites respectively) is verified for its use in forecasting the possible occurrence of sea storms over the Adriatic Sea. For this purpose, verification is performed over a set of coastal stations. A novel summary-plot of the scores derived from the contingency table is here presented, denominated "Performance Rose". In a wind-rose plot, the observed wind frequency is represented subdivided into the usual 8 sectors. In additions to this, the scores for each sector (i.e. for each direction) are also represented: the Frequency Bias by the colour of the wind-rose "petal", the Probability of Detection, the Threat Score and the Success Ratio by symbols on the appropriate scale (from 0 to 1 for all of them) plotted in the radial axes. The verification is performed on each station point separately to take into account the dependence of the low-level wind on the topography. It is also possible to aggregate the forecast issued by the model within a radius as predictor of the wind on the same point, in order to increase the statistical robustness of the estimate. The high temporal variability of the parameter can be also taken into account by performing the verification on appropriate time intervals, considering the prevailing wind speed and direction in each interval. The performance of this verification methodology is shown on selected periods, including different kinds of weather events, focusing on the information which can be derived from this approach.

USR VAL-3: User-oriented evaluation of fire spread predictions

AUTHORS: Beth Ebert[1], Chris Bridge[2], Nathan Faggian[2], Barbara Brown[3], Paul Fox-Hughes[1], Howard Jacobs[2], Stuart Matthews[4], Greg Esnouf[5]

Presenting author e-mail: beth.ebert@bom.gov.au; Affiliations: [1]Research and Development Branch, Bureau of Meteorology, Australia, [2]Weather Forecasting Branch, Bureau of Meteorology, Australia, [3]Research Applications Laboratory, National Center for Atmospheric Research, USA, [4]New South Wales Rural Fire Service, Australia, [5]Australasian Fire and Emergency Service Authorities Council, Australi

ABSTRACT: Fire and land management agencies in Australia and elsewhere increasingly rely on physical and empirical fire spread models to provide decision support and risk modelling capability for incident management and planned fuel reduction activities. However, it is widely acknowledged that fire spread modelling is subject to numerous shortcomings including limitations in the underlying equations that relate the rate of fire spread to the fuel state, weather conditions and topography. Moreover, there are significant uncertainties in the ignition, fuel and weather inputs required to drive the models. Fire behaviour analysts (FBANs), the main users of fire models, need information to help decide which model(s) to use and how much confidence to place in their output. The Bureau of Meteorology is collaborating with regional fire services, universities, and NCAR to develop an evaluation framework to help FBANs make informed choices on fire spread models. A generic software framework enables different fire spread models to be installed and run in a consistent way. We gathered reliable observations of burned areas for a number of wildfire events from around Australia to verify model predictions of fire area, spread rate and direction, which are amongst the variables of greatest interest to FBANs. We ran the models using the best operational weather and ignition inputs, as well as inputs with realistic levels of uncertainty applied. An early finding was the large sensitivity of some models to small deviations in ignition location, which led the NSW Rural Fire Service to change the way they ran the fire model in operations. Communicating the accuracy and the sensitivity of the models to users in a concise yet understandable way has been challenging given the multi-dimensional nature of the evaluation. We have found a way to display accuracy and sensitivity together using an enhanced Hinton (matrix) diagram that portrays accuracy using color and sensitivity using boxes of different size. Categorical performance diagrams are used to inter-compare the performance of different models. Continuous consultation with the end-users has been a critical component of this effort, producing an evaluation which will maximize value for decision makers.

USR VAL-4: Develop a user oriented forecast verification metric

AUTHORS: Wang, William (Xiangdong); Watkins, Andrew; Jones, David.

Presenting author e-mail: William.Wang@hotmail.com; Affiliation (all authors): Australian Bureau of Meteorology, Melbourne, Australia

ABSTRACT: Climate Information Services (CIS) of the Bureau of Meteorology has been issuing the national seasonal ‚above median' probability outlooks for rainfall and maximum and minimum temperature public every month (http://www.bom.gov.au/climate/outlooks). Verification of these outlooks is considered a key aspect of the service. The Bureau calculates a number of verification metrics for these probability forecasts, including LEPS, Brier Score and ROC score. However, these metrics have been found to be difficult to communicate and interpret. Instead, percent-consistent (PC) is used to show the skill of forecasts. But a critical issue of PC is it doesn't consider the magnitude of the actual outcome. A skilful forecast model is expected to capture the significant and predictable part of the variation, and such capability is not assessed by PC or the other metrics mentioned above. To keep the simplicity of PC but tackle the issue and improve the effectiveness of this metric, the so called "weighted percent consistency (WPC)" is designed in this paper. The paper has then approved that observed anomaly above climatological median is the most appropriate weighting element among the general statistics, such as deciles, percentage (for rainfall only), anomaly from climatological mean etc. The paper also demonstrated that WPC satisfies the general rules of skill metric, and it is robust, intuitive and easy to calculate. Reference skill is 0.5, below 0.5 represents no skill, and the skill for perfect forecast is 1.

USR VAL-5: The final approach ? Verification of Terminal Aerodrome Forecasts

AUTHORS: Sharpe, Michael; Bysouth, Clare.

Presenting author e-mail: michael.sharpe@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: Terminal Aerodrome Forecasts (TAFs) are a widely accepted international form of aviation forecast used for airport and flight planning procedures at all major airports; these abbreviated text forecasts contain probabilistic, deterministic and temporal uncertainty terms which make verification challenging. This presentation outlines a new verification approach which adopts the WMO definition of each TAF abbreviation, matching the observations to the forecast as closely as these definitions allow. Then the performance is measured using a novel multi-category reliability table approach; an analysis of the visibility component to TAFs is used to demonstrate that this methodology performs well when compared with existing verification approaches in a variety of different test cases which have been chosen to illustrate scenarios that are important to forecast correctly. (Additional comment: we have also submitted this for the verification challenge.)

USR VAL-6: Displaying the Verification Results of Terminal Aerodrome Forecasts for Thunderstorms and Visibility

AUTHORS: Jurkovic, Jadran; Pasaric, Zoran; Kos, Igor.

Presenting author e-mail: jadran.jurkovic@crocontrol.hr; Affiliation (all authors): Croatia Control Ltd, Velika Gorica, Croatia

ABSTRACT: As part of a quality management system, ICAO Annex 3 requires verification of forecasts. Terminal Aerodrome Forecast (TAF) is important in decision making during the flight planning process. Verification procedures are not standardized but one of the leading approaches is proposed by Mahringer (2008). It is based on verifying the best and worst conditions for each hour, between observed and forecast states. Data used for verification are TAF forecasts and observed reports (METAR) which are issued half-hourly. The occurrence of a verified event is rather rare owing to three reasons; because the forecast is for a given point, the verification time period is just one hour and aviation requirements usually refer to high impact weather. At Croatian airports the occurrence of thunderstorms is about 1-2% in terms of hours yearly, hence thunderstorm can be regarded as a rare event. Therefore, for comparing results the diagram with bias and tetrachoric correlation coefficient (TCC) is used. Together with climatological probability, this triplet fully describes contingency tables (Juras and Pasaric, 2006). Moreover, it is suggested that TCC is a good measure of association of a contingency table for rare events. When verifying visibility, a multi categorical contingency table with class limits arising from criteria for groups of changes in a TAF forecast is calculated. Consequently, instead of TCC, the polychoric correlation coefficient (PCC) is used. Bases on this approach, several applied diag rams which reveal verification results to users are shown.

META-1 KEY NOTE: Generating and verifying probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations

AUTHORS: Ferro, Chris A.T..

Presenting author e-mail: C.A.T.Ferro@exeter.co.uk; Affiliation (all authors): University of Exeter , Exeter, UK

ABSTRACT: The performance of forecasts is multifaceted, but it is often useful to summarise performance with a single measure, and scoring rules are a good way to do this. We revise the notion of a proper scoring rule and other, related types of scoring rule. We discuss the interpretation of scoring rules and how they measure different aspects of forecast performance, such as calibration and sharpness. We also describe how to make inferences using scoring rules, for example to compare the performance of two forecasters.

META-2: Proper and equitable scores: a resolved dilemma

AUTHORS: Ben Bouallegue, Zied.

Presenting author e-mail: zied.benbouallegue@ecmwf.int; Affiliation (all authors): ECMWF, Reading, UK

ABSTRACT: Propriety and equitability are desirable properties of scoring rules, that is functions dedicated to the evaluation of probabilistic forecasts. On one side, the use of proper scores is required in order not to encourage hedging, so the expected verification results are optimised when the forecaster expressed his/her true belief. On the other side, the use of equitable scores ensures that all non skilful forecasts have the same expected verification results. However, propriety and equitability have been considered so far as incompatible properties. In a general framework for the definition of user oriented scores, it is shown that a family of scores that are simultaneously proper and equitable can be defined by making a simple assumption about the relationship between users' risk aversion and frequency of occurrence the events of interest. These so-called diagonal scores are compared and related to common verification measures to illustrate their astonishing statistical properties. Moreover, an interpretation of the diagonal scores in terms of forecast value for vulnerable users is proposed for a simplified communication of the derived verification results.

META-3: The generalized discrimination score: connections, corrections and potential applications

AUTHORS: Harbord, Roger.

Presenting author e-mail: roger.harbord@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: The generalized discrimination score was introduced by Simon Mason and Andreas Weigel in 2009 (Mon. Weather Rev. 137:331-349) as a non-parametric measure of forecast discrimination with a reasonably straightforward interpretation, applicable to all scales of measurement of observations and to both deterministic and probabilistic forecasts. They extended its definition to ensemble forecasts in 2011 (Mon. Weather Rev. 139:3069-3074). We briefly summarise these papers and clarify connections between this score and similar metrics previously published in the statistical literature: for deterministic forecasts it is a simple rescaling of Somers' D, which may be thought of as a generalization of the area under the receiver operating characteristic (ROC) curve beyond dichotomous outcomes. We unify the many formulations provided by Mason and Weigel for the various combinations of observation and forecast type. We find that they make an unwarranted assumption of independence in their formulation of the score for probabilistic forecasts, including ensemble forecasts. Although removing this assumption appears to make little difference to the score in practice, it suggests a simpler and faster algorithm for ensemble forecasts that scales better with ensemble size. We apply the score to some example deterministic, ensemble and percentile forecasts to illustrate its potential use in seamless verification of the post-processing chain for assessing the gain or loss in discrimination provided by each step of the chain.

META-4: Decomposition and Attribution of Forecast Errors

AUTHORS: Yang, Fanglin.

Presenting author e-mail: Fanglin.Yang@noaa.gov; Affiliation (all authors): Environmental Modeling Center, National Centers for Environmental Prediction, College Park, Maryland, USA

ABSTRACT: Root-Mean-Square Error (RMSE) has long been used as a performance metric for evaluating climate and weather forecast models. In this presentation it will be shown analytically that RMSE at times misrepresents model performance. A decomposition technique is proposed to describe more precisely RMSE distributions. Conventional RMSE can be decomposed into Error of Mean Difference (E_m) and Error of Patter Variation (E_p). E_p is unbiased and can be used as an objective measure of model performance only if the anomalous pattern correlation (R) between forecast and analysis approaches to one. If R is small, E_p is biased and favors smoother forecasts that have smaller variances. E_p normalized by analysis variance is still biased and favors forecasts with smaller variance if anomalous pattern correlation is not perfect. A comprehensive evaluation of model errors should include Anomalous Pattern Correlation, Ratio of Forecast Variance to Analysis Variance, Error of Mean Difference, and Error of Pattern Variation. NCEP Global Forecast Systems with different configurations will be used to demonstrate the decomposition technique, which will be applied to both scalar variables and vector winds. At the end, the advantage and limitation of using data assimilation analysis increments to diagnose model biases will be discussed.

META-5: Murphy diagrams

AUTHORS: Jordan, Alexander; Ehm, Werner; Gneiting, Tilmann; Krüger, Fabian.

Presenting author e-mail: alexander.jordan@h-its.org; Affiliation (all authors): Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

ABSTRACT: When is a forecast system preferable over another, irrespective of the (consistent or proper) scoring function being used? As it turns out, in the case of probability forecasts for a binary event, and also in the cases of expectation (mean) and quantile forecasts for real-valued variables, the consistent scoring functions admit mixture representations in terms of easily interpretable elementary members. The elementary scoring functions form a linearly parameterized family, where the linear parameter corresponds to a threshold in a decision-making problem. Therefore, we can plot the empirical scores of competing forecast systems as a function of relevant threshold values. We call this type of display a Murphy diagram, in honor of Allan H. Murphy. Murphy diagrams offer simple checks of whether a forecast dominates another, in the sense that it is preferable under any consistent scoring function, and permit detailed comparisons of the relative merits of competing forecast systems. In my presentation, I will sketch the underlying theory and illustrate the use of Murphy diagrams in meteorological and economic case studies.

META-6: Verification of Multi-Valued Forecasts

AUTHORS: Bullock, Tim; Shum, Nelson.

Presenting author e-mail: Timothy.Bullock@canada.ca; Affiliation (all authors): Meteorological Service of Canada, Halifax, Canada

ABSTRACT: Some weather forecasts predict physical quantities which assume multiple values during a time interval and/or over a region of interest. Such forecasts must necessarily be multiple-valued as well, otherwise they could not accurately predict those quantities. In other cases, forecast values are organized into categories which are not mutually-exclusive. In still others, there is a large discrepancy between the precision of forecast and observed categories. In each of these instances, application of traditional, single-valued verification methods (i.e. comparing one forecast value with one observed value) may entail significant loss of information, which can compromise the validity or usefulness of conclusions obtained from the analysis of performance. By relaxing the requirement that forecasts and observations be single-valued, it is possible to preserve the informati on content of both forecasts and observations, thereby offering additional insights into the performance of the aforementioned forecast systems. Insofar as a multi-valued forecast system could be considered as a generalization of a single-valued system, the verification metrics for a multi-valued forecast system might also be viewed as generalizations of their single-valued counterparts. A selection of metrics is proposed for multi-valued verification (MVV). Application of the MVV approach is then demonstrated for forecasts of a few weather variables from public and marine forecast bulletins issued routinely by the Meteorological Service of Canada.

PROB ENS-1: Ensemble verification: Old scores, new perspectives

AUTHORS: Wahl, Sabrina; Friederichs, Petra; Keller, Jan D..

Presenting author e-mail: wahl@uni-bonn.de; Affiliation (all authors): Hans-Ertle-Centre for Weather Research / University of Bonn, Bonn, Germany

ABSTRACT: The recently derived decomposition of the quantile score into uncertainty, resolution and reliability (Bentzien and Friederichs, 2014) can be generalized to a decomposition of the continuous ranked probability score (CRPS). This decomposition reveals detailed insights into ensemble performance in terms of calibration and information content. We utilize this framework to investigate the performance of raw ensemble forecasts, i.e. without further translation of the ensemble members into probabilistic predictions. Such an evaluation is of particular interest for developers and designers of ensemble prediction systems. We further present a novel graphical exploration of the forecast attributes.

PROB ENS-2: Increasing the usability of analysis rank histograms

AUTHORS: Keller, Jan; Wahl, Sabrina.

Presenting author e-mail: jkeller@uni-bonn.de; Affiliation (all authors): Hans-Ertel-Centre for Weather Research / Deutscher Wetterdienst, Offenbach, Germany

ABSTRACT: An analysis rank histogram (ARH) is a popular tool for evaluating probabilistic data sets comprising multiple realizations of an observed parameter, i.e., an ensemble. By looking at the ARH one can determine the calibration of an ensemble, i.e., how well does the ensemble represent the observed variability. A flat ARH is usually a sign for the ensemble to be reliable while the often encountered U-shaped histogram indicates un under-dispersiveness of the ensemble, i.e., the variability is too small. The evaluation of ensemble performance using ARHs is generally done by averaging over large data sets of forecasts and observations for multiple locations and time steps. Depending on the application, a stratification of calibration with respect to geographic regions and/or specific time periods may be desirable. However, if one is interested in the spatial distribution or the temporal evolution of ensemble calibration, it would be difficult to visually assess the information of a lot of ARH plots or to put them into a journal publication. The beta score provides a simple approach to tackle this problem. Using a parametric approximation of the ARH to a beta distribution, the score allows for the description of the histogram shape with a single value. We show examples of ARHs and the corresponding beta scores to confirm the validity of the approximation. We than compare reliability assessments for different ensemble forecasts using beta score time series as well as map plots. We will further show results for a covariance evaluation of these forecasts using ARH-based beta scores.

PROB ENS-3: Another look at spread and skill

AUTHORS: Johansson, Åke.

Presenting author e-mail: Ake.Johansson@smhi.se; Affiliation (all authors): Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

ABSTRACT: The relationship between spread and skill is an important indicator of the quality of an ensemble prediction system. Arguments are given that suggests that an alternative to the commonly used spread-skill condition is needed to properly address the true relationship between spread and skill. Such a relationship is derived and given the name U.UI spread-skill condition. Its properties are described both analytically and geometrically and the relationship to the commonly used condition is demonstrated and discussed. It is argued that it would provide not only a more appropriate and sharper tool to assess the spread-skill relationship, but also that it indicates how to improve upon the present design of an EPS. In particular, the presently prevalent practice of centering the perturbed ensemble members around the control analysis should be abandoned in favor of a methodology that constructs equally likely ensemble members with the same quality as the control member. A recent experiment with a limited area ensemble prediction system is used to quantify the differences in perceived quality that is obtained by employing the two different spread-skill conditions. The differences are found to be substantial.

PROB ENS-4: The development of probabilistic forecast verification across the scales

AUTHORS: Zhu, Yuejian; Melhauser, Christopher; Li, Wei.

Presenting author e-mail: Yuejian.Zhu@noaa.gov; Affiliation (all authors): Environmental Modeling Center, NCEP/NWS/NOAA, College Park, MD 20740, U. S. A

ABSTRACT: In past years, ensemble based probabilistic forecast verification has been developed to evaluate ensemble forecast performance. The methodologies, such as Rank Histogram (RH), Brier Score (BS), Ranked Probability Score (RPS), Continued Ranked Probability Score (CRPS), Relative Operational Characteristics Score (ROCS), Economic Value (EV) and others, have been used widely in weather prediction and other application. In this study, we will focus on multi-categories with probabilistic forecast that is different from a binary forecast either occurs, or does not occur, to measure the quality of ensemble based probabilistic forecast. An investigation includes the performance difference for different spatial and time scales. In order to fully understand probabilistic forecast performance and skills, we analyses the forecast distributions with large/small domain average, short/long-term average, and the climatology from different records. A forecast skill highly depends on the categories we have selected, and the references we have used. For short term and weather forecast, we usually predict element instantly, such as temperature; or short-period, such as precipitation for one-hour, 3-hours, 24-hours accumulation and etc. Starting from week-2 and beyond, we issue forecast anomaly for weekly, two-weekly, monthly, and seasonal average based on defined categories, such as tercile forecast, above/below normal forecast. In this work, we will use NCEP global experimental ensemble 35-day forecast to demonstrate a forecast skill changes with not only forecast lead-time, but also a category/reference we have used. This will guide us to understand forecast skills across scales.

PROB ENS-5: Generating and verifying probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations

AUTHORS: Schwartz, Craig; Sobash, Ryan.

Presenting author e-mail: schwartz@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research (NCAR), Boulder, United States

ABSTRACT: “Neighborhood approaches“ have been used in two primary ways to post-process and verify high-resolution ensemble output. While the two methods appear deceptively similar, they in fact define events over different spatial scales and yield fields with different interpretations: the first produces probabilities interpreted as likelihood of event occurrence at the grid-scale, while the second produces probabilities of event occurrence over spatial scales larger than the grid-scale. Unfortunately, some studies have confused the two methods, while others did not acknowledge multiple possibilities of neighborhood approach application and simply stated, “a neighborhood approach was applied“ without supporting details. Thus, this presentation overviews applications of neighborhood approaches to convection-allowing ensembles in hopes of clarifying the two methods and their different event definitions. Additionally, it will be shown that the two approaches can yield statistically significantly different objective conclusions about model performance, underscoring the critical need for thorough descriptions of how neighborhood approaches are implemented and events are defined. Finally, some recommendations for application of neighborhood approaches to convection-allowing ensemble forecasts of high impact weather will be provided.

PROB ENS-6: Benefit of an ensemble over a deterministic forecast at km-scale

AUTHORS: Mittermaier, Marion.

Presenting author e-mail: marion.mittermaier@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: What is the benefit of a near-convection-resolving ensemble over a near-convection-resolving deterministic forecast? Ensemble and deterministic Numerical Weather Prediction(NWP) forecasts can be compared using a probabilistic verification framework. Three years of raw forecasts from the Met Office Unified Model (UM) 12- member 2.2 km MOGREPS-UK ensemble and 1.5 km UKV deterministic configuration were compared, utilising a range of forecast neighbourhood sizes centred on surface synoptic observing site locations. Six surface variables were evaluated: temperature, 10 m wind speed, visibility, cloud base height, total cloud amount and hourly precipitation. Deterministic forecasts benefit more from the application of neighbourhoods, though ensemble forecast skill can also be improved. This confirms that whilst neighbourhoods can enhance skill by sampling more of the forecast, a single deterministic model state in time can not provide the variability, especially at the kilometer-scale, where rapid error growth acts to limit local predictability. Ensembles are able to account for uncertainty at larger, synoptic scales. The results show that the rate of decrease in skill with lead time is greater for the deterministic UKV. The concept of a skill differential is introduced to find the smallest neighbourhood size at which the deterministic and ensemble scores are comparable. This was found to be 3 x 3 (6.6 km) for the ensemble, and 11 x 11 (16.5 km) for deterministic. Comparable scores are between 2-40% higher for the ensemble, depending on the variable. Naively, this would suggest that an extra 10 km in spatial accuracy are gained by using a km-scale ensemble.

OPS SOFT-1: The difficulty of verifying small improvements in forecast quality

AUTHORS: Geer, Alan.

Presenting author e-mail: alan.geer@ecmwf.int; Affiliation (all authors): ECMWF, Reading, UK

ABSTRACT: To verify an upgrade to an operational forecasting system it is typical to compare an experiment, in which something relatively small has changed, to a control which is already capable of making forecasts of very high quality. Drawing valid conclusions from these experiments is hard. Chaotic variability in forecast quality makes it difficult to achieve statistical significance, to the extent it could require years of experimentation to identify a small improvement (Type II error). There is also great danger of finding false positives in the significance testing (Type I error) unless this accounts for autocorrelation in the forecasts and for multiple comparisons. Further, all this assumes that the verification itself is robust. However, in the early forecast range, and at any forecast range in the tropics and stratosphere, this is rarely true. For example the verification reference is usually correlated with the forecasts, even if the verification is done against observations: most observations go into the analysis and hence influence the forecast. This can lead to apparently significant changes in forecast scores that might just come from changes in the model‘s climatology or in the size of data assimilation increments. This presentation summarises recent experience in the research department at ECMWF, giving an overview of the problems and recent developments that try to address them, such as improved techniques for statistical significance testing; verification on lat-lon and lat-pressure grids; and the use of verification based on both observations and analysis.

OPS SOFT-2: Methods for Identifying Small Improvements in Forecasts

AUTHORS: Weatherhead, Elizabeth; Jensen, Tara; Brown, Barb.

Presenting author e-mail: betsy.weatherhead@colorado.edu; Affiliation (all authors): University of Colorado, Boulder, Colorado, USA

ABSTRACT: Weather forecasts have been improving over the last seven decades through a series of small improvements in model development, parameterizations, physics, data assimilation, new observations and computational advances. Not all proposed changes result in improvements. Testing and identifying when there has been a true improvement depends on a number of factors. This presentation will review some of the aspects of model design and statistical significance that can help identify true improvements that can lead to long-term forecast improvements. Factors that are important are sample size, representativeness of weather regimes, autocorrelation, and deviations from Gaussian behavior. The presentation will focus on the statistical tests that can help with decision support. Examples from the US Next Generation Global Prediction System (NGGPS) efforts will be used.

SPATIAL-1 KEY NOTE: Verification of user-relevant contours with the Integrated Ice Edge Error and the Spatial Probability Score

AUTHORS: Goessling, Helge.

Presenting author e-mail: helge.goessling@awi.de; Affiliation (all authors): Alfred Wegener Insitute, Bremerhaven, Germany

ABSTRACT: The pan-Arctic sea-ice extent (SIE) is a prominent example of a common forecasting target of very little user relevance: a forecasting system with excellent skill in predicting the SIE can still place the ice in wrong locations, preventing vessels from reliable planning of routes through ice-infested waters. A simple verification metric that overcomes the problem of compensating errors is the Integrated Ice Edge Error (IIEE), defined as the total area of mismatch between the predicted and observed ice cover. The decomposition into Absolute Extent Error and Misplacement Error components allows further specification of forecast skill. Moreover, the Spatial Probability Score (SPS), defined as the spatial integral of local (Half) Brier Scores, can be regarded (when applied to the ice edge) as a probabilistic generalization of the IIEE. The SPS can also be considered the spatial analog of the Continuous Ranked Probability Score. In my presentation I will demonstrate these new verification metrics based on idealised (perfect-model) ensemble forecasts of the Arctic sea-ice edge, revealing in particular that the SPS responds properly to ensemble size, spread, and bias. I will also put initialised forecasts in context with climatological and persistence forecasts, allowing to assess the potential predictability of the Arctic sea-ice edge from a probabilistic viewpoint. Finally, I will present preliminary results for real-world sea-ice forecasts based on these verification metrics.

SPATIAL-2: Verification of sea-ice prediction by using distance measures

AUTHORS: Casati, Barbara; Lemieux, Jean-Francois; Smith, Greg; Pestieau, Paul.

Presenting author e-mail: barbara.casati@canada.ca; Affiliation (all authors): Meteorological Research Division, Environment and Climate Change Canada, Dorval, Canada

ABSTRACT: Sea-ice is characterized by a coherent spatial structure, with sharp discontinuities and linear features (e.g. leads and ridges), the presence of spatial features, and a multi-scale spatial structure (e.g. agglomerates of floes of different sizes). Traditional point-by-point verification approaches do not account for this complex spatial structure and the intrinsic spatial correlation existing between nearby grid-points. This leads to issues (such as double penalties), and an overall limited diagnostic power (e.g. traditional scores are insensitive to distance errors). This work explores the use of binary image distance measures of the Hausdorff and Baddeley family for the verification of sea-ice extent and sea-ice edge. The metrics are illustrated for the Canadian Regional Ice Ocean Prediction System evaluated against the Ice Mapping System analysis. The distance measures account for the sea-ice coherent spatial structure, and are sensitive to the overlapping and similarities in shape of observed and predicted sea-ice extent: they reveal to be a robust and suitable set of verification measures, complementary to the traditional categorical scores. Moreover, these measures can provide distance errors, e.g. of observed versus predicted sea-ice edge, in physical terms (i.e. km), thereby being informative and meaningful for user-relevant applications.

SPATIAL-3: Comparing distance methods for spatial verification

AUTHORS: Gilleland, Eric.

Presenting author e-mail: EricG@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research, Boulder, Colorado, U.S.A.

ABSTRACT: When performing feature-based forecast verification, it often becomes necessary to calculate a distance, in terms of location, between two features. Other times, it can be of interest to compare entire fields in this way. Some common distance methods are compared, including: centroid distance, Baddeley‘s Delta metric, and the mean error distance. Pros and cons of the measures are summarized to give a sense of which can be used for specific instances.

SPATIAL-4: Estimating the displacement in precipitation forecasts using the Fractions Skill Score

AUTHORS: Skok, Gregor; Roberts, Nigel.

Presenting author e-mail: gregor.skok@fmf.uni-lj.si; Affiliation (all authors): Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia, Ljubljana, Slovenia

ABSTRACT: The Fractions Skill Score (FSS) is a popular spatial verification metric commonly used for precipitation verification. In this study we focus on analysing the ability of FSS to provide meaningful information about the displacement between precipitation in one field compared to another. Towards this goal a number of relevant idealized cases were analysed that showed that the FSS can indeed be used to determine displacement in a meaningful way. It was found that the displacement provided by the FSS is directly related to the true displacements of precipitation but with larger precipitation area having a much larger influence. Overall, the displacement provided via the FSS tends to correspond quite well to the average distance to the closest neighbouring raining area (assuming the areas are of similar size). The user should alw ays use a frequency threshold and make sure to use the overlap-adjusted variant of the FSS displacement. The user should also avoid using FSS displacement in cases of large precipitation bias.

SPATIAL-5: A multi-model, multi-analysis study to asses the capability of CRA analysis for QPF spatial verification in the MesoVICT framework

AUTHORS: Mariani, Stefano; Casaioli, Marco.

Presenting author e-mail: stefano.mariani@isprambiente.it; Affiliation (all authors): ISPRA - Institute for Environmental Protection and Research, Rome, Italy

ABSTRACT: As contribution to the Mesoscale Verification Inter-Comparison over Complex Terrain (MesoVICT) project, the present study investigates the pros and cons in applying the contiguous rain area (CRA) analysis to verify high-resolution quantitative precipitation forecasts (QPFs) over a suitable Central Europe region. This region is characterized by complex terrain due to the simultaneous presence of the Alps (i.e., complex orography) and the Mediterranean Sea (i.e., lack of observations, coastlines). The spatial verification method under investigation assesses (and qualifies) the displacement and/or distortion required to match the QPF features to the related observation features. Complex patter-matching criteria and additional constraints result to be necessary to discriminate between feasible pattern matches and unrealistic ones, especially when considering such a domain where the rainfall entity could not be completely observed (e.g., over the sea) or forecast (i.e., the rainfall feature is forecast, at least partly, outside the verification domain). Moreover, the CRA results are found to be sensitive to model native resolution, especially when observation spacing is seemingly coarser than model grid. The capability of the CRA analysis is demonstrated by considering all the six MesoVICT cases, which occurred in 2007 during the MAP D-PHASE Operations Period (DOP), with a major focus on case #1 (20-22/6/2007 the core event) and case #3 (25-28/9/2007), plus one additional high impact DOP event (22-25/11/2007). Two datasets of gridded rainfall fields are considered: the 8-km observational fields obtained by using the Vienna Enhanced Resolution analysis (VERA) scheme; and the 10-km observational fields obtained by using a two-pass Barnes scheme. For comparison over the VERA grid, QPFs are from the Swiss model COSMO-2 and the Canadian high-resolution model GEM-LAM; whereas for comparison against the Barnes analyses, QPF are from the ISPRA hydrostatic BOLAM model (in two different configurations, with grid spacing of 0.07 and 0.1°) and the ISPRA high-resolution convection-permitting MOLOCH. Results are shown for discussion during the workshop.

SPATIAL-6: Evaluation of Extra-Tropical Cyclones Using a Feature-Relative Method

AUTHORS: Jensen, Tara; Halley Gotway, John; Kucera, Paul; Colle, Brian.

Presenting author e-mail: jensen@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research/Research Applications Lab, Boulder, Colorado, USA

ABSTRACT: Many of the current verification metrics used in operations are grid point relative (e.g., regional), and averaging over a fixed geographic region will smooth many important processes and significant weather features around the cyclone. Some verification metrics are gathered for cyclones, but they are usually limited to central pressure, displacement, and cyclone density errors. In order to better understand operational model and ensemble performance and processes that may be leading to any systematic errors of significant weather around the cyclone, a cyclone-relative verification approach is needed. Knowledge of the performance of cyclone-relative features, such as jets, precipitation bands, and fronts, can also help develop confidence metrics and tools for forecasters about the skill in predicting those features at various lead times from 0 to 10 days. A feature-relative approach was added to the unified modeling framework built on the Model Evaluation Tools (MET). MET is a state-of-the-science verification package supported to the community through the Developmental Testbed Center. Standard measures, such as continuous statistics may be computed in this cyclone relative framework. The errors associated with the cyclones can then related to the various important physical processes (moisture flux, stability, strength of upper-level PV anomaly and jet, surface fluxes, and precipitation), and the results will be separated by different stages of the cyclone (genesis, mature, and decay), cyclone intensity, and large-scale flow regime. This approach will be transferred to the U.S. National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Prediction (NCEP) to help model developers and forecasters better understand the origin of cyclone biases (e.g., hypothesized under-deepening of surface cyclones in the G FS in the medium range), and the ensemble performance for the significant weather around the cyclone. This presentation will describe the feature-relative methodology and provide examples from the prototype system.

SPATIAL-7: Methods for Evaluation of Cloud Predictions

AUTHORS: Brown, Barbara; Jensen, Tara; Gilleland, Eric; Halley Gotway, John; Fowler, Tressa; Bullock, Randy.

Presenting author e-mail: bgb@ucar.edu; Affiliation (all authors): NCAR, Boulder, United States

ABSTRACT: Cloud predictions are very important for many applications, yet verification of cloud forecasts remains challenging. In particular, evaluation of the ability of gridded cloud predictions replicate spatial features is not straightforward due to the complex nature of cloud fields (e.g., multiple layers) and the limitations associated with cloud observations. Some of the new spatial verification methods (e.g., object-based, neighborhood, scale separation, image warping), which have primarily been applied to numerical weather prediction (NWP) model forecasts of precipitation, show promise as alternative methods for evaluating cloud predictions. Although these approaches are suitable for application to any variable characterized by coherent spatial features, they have not been widely applied to other fields. In a project collaboration between NCAR and the U.S. Air Force 557th Weather Wing, a number of approaches have been examined for evaluation of gridded cloud predictions. The forecasts included in the study consist of advection products as well as predictions from NWP models. The verification approaches examined include the Method for Object-based Diagnostic Evaluation (MODE); several distance measures; the Structure, Amplitude, and Location (SAL) approach; and image warping. Results of application of these different approaches to a set of advection and NWP-based cloud amount predictions will be presented and compared; positive and negative aspects associated with application of each method will be discussed.

SPATIAL-8: Grid-based spatial verifications versus point-based verifications with the application to precipitation and clouds

AUTHORS: Jenkner, Johannes; Oberthaler, Andrew; Mayer, Dieter; Spatzierer, Manfred.

Presenting author e-mail: jjenkner@ubimet.com; Affiliation (all authors): UBIMET GmbH, Vienna, Austria

ABSTRACT: In operational verification, reference data can be of different type, but most commonly comprise (a) discrete station observations or (b) gridded observational analyses. In this study, gridded NWP model forecasts for precipitation and cloud cover are verified over the entire D-A-CH region and 40 selective sub-regions using both reference data types. For (a), several hundred stations with hourly observations are used as verification locations onto which the model forecasts are interpolated to. For (b), the Vienna Enhanced Resolution Analysis (VERA) is applied and spatially consistent fields of precipitation and cloud cover are constructed by blending local observations with remote sensing data. In such a way, VERA serves as a high-quality gridded reference for an area-wide verification. Several case studies in convective as well as stratiform situations are presented and different model forecasts are compared to each other. To gain insights into the modeling of rainfall intensities and cloud amounts, categorical statistics based on 2x2 contingency tables are applied to all data. Here, potential differences in the results for (a) and (b) are highlighted and interpreted by using domain knowledge on the verification region. To allow for an object-based perspective, the SAL measure is applied to the gridded data of (b) and modeled fields are examined with respect to the structure, amplitude and location of rainfall and cloud objects. Finally, the verified models are rated using the output from both the categorical measures and the SAL and it is shown that a comprehensive evaluation is only possible by applying different verification measures at the same time.

OBS UNC-1: A Unified Approach to Verification

AUTHORS: Jensen, Tara; Strong, Bonny; Stajner, Ivanka; Yang, Fanglin; White, Glenn; Brown, Barbara.

Presenting author e-mail: jensen@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research/Research Applications Lab, Boulder, Colorado, USA

ABSTRACT: The need to simplify the U.S. National Center for Environmental Prediction (NCEP) production suite and accompanying infrastructure was identified both by several committees and working groups in 2015. The goal of the unification is to facilitate evidence based decision making with regards to model and forecast upgrades and to allow for a broader community approach to model improvements. The initial focus of unification is on the global and regional numerical weather prediction scales but is expected to eventually extend to climate scales as well. The unification is being performed through a combined effort between the Developmental Testbed Center and the National Oceanic and Atmospheric Administration (NOAA) Next Generation Global Prediction System (NGGPS) program. The NGGPS Verification and Validation team reviewed many different options for building the capabilities needed to support NGGPS through a unified software package. They decided to base the unified system on the existing Model Evaluation Tools (MET) package supported by DTC, along with its associated interactive web-based aggregation and visualization package METViewer. The unified system is called MET+, which reflects the use of MET and METViewer at the core with a set of flexible python-based scripts to wrap the system. Phase 1 of this effort is nearing completion and includes methodologies taken from many aspects of the verification community. This presentation will provide an overview of the system and examples from the initial system.

OBS UNC-2: Evaluating representativeness errors in verification against Arctic surface observations

AUTHORS: Haiden, Thomas.

Presenting author e-mail: thomas.haiden@ecmwf.int; Affiliation (all authors): ECMWF, Reading, United Kingdom

ABSTRACT: In the verification of numerical weather prediction models against point observations, the total forecast error can be separated into a grid-scale forecast error and a representativeness error, which arises from the scale mismatch between the model grid and the effective spatial representativeness of the observation. Evaluated over a certain time period, representativeness depends on the aggregation time interval (near-instantaneous values, daily or monthly averages), the kind of metric used (such as, for example, bias, root-mean-square error, or correlation), and whether actual values or deviations from climatology (anomalies) are considered. By comparing short-range NWP forecasts and analyses of surface downward longwave (DLR) radiation against in-situ surface observations, we try to quantify the representativeness error . DLR is one of the key parameters influencing inter-annual sea ice variability and depends on the vertical profiles of atmospheric temperature and moisture, as well as cloud conditions. The International Arctic Systems for Observing the Atmosphere (IASOA) observatories at which the measurements are made, are mostly located at coastal sites and have the potential of representing both continental and oceanic (open water and sea ice) conditions. High-resolution operational analyses and ERA-Interim re-analyses are used to assess the spatial representativeness of the observations at different time-scales. It is found that at distances exceeding a few 100 km, correlations derived within re-analyses are similar to correlations between point observations and analyses, thereby offering a way of evaluating the potential benefit of future additional observation sites. On the local scale, DLR observations with high temporal resolution (1 min) are used to estimate sub-grid scale spatial variabi lity invoking Taylor?s hypothesis. Applicability of the methodology to other surface parameters such as 2 m temperature, 10 m wind speed, and precipitation is discussed.

OBS UNC-3: Spatial verification of ensemble forecasts with SAL and the role of observation uncertainty

AUTHORS: Radanovics, Sabine.

Presenting author e-mail: sabine.radanovics@lsce.ipsl.fr; Affiliation (all authors): LSCE, Gif-sur-Yvette, France

ABSTRACT: With the increased resolution of modern ensemble forecasts, their gridpointwise verification now faces the same issues the verification of deterministic forecasts encountered a decade ago, for example the double penalty problem. Therefore the spatial verification methods that were designed to solve these issues have to be adapted for ensemble forecasts. First, I present an ensemble version of the SAL spatial verification method. Then the ensemble version and the original deterministic SAL are compared using precipitation data from the core case of the MesoVICT project, including deterministic and ensemble forecasts (COSMO2, CLEPS) and deterministic and ensemble analysis (VERA). The comparison results show that the ensemble S and A components are close to the median of the score distribution resulting from calculating the deterministic SAL for every ensemble forecast member-ensemble analysis member pair, while the ensemble L tends to be smaller than the median of deterministic L scores. The ensemble SAL can diagnose the structure and amplitude biases introduced in the VERA ensemble analysis due to the way the analysis ensemble was constructed. These biases are, for very short forecast lead times of less than one day, larger than the systematic differences between the two forecast models. The ensemble SAL has the advantage of being faster to calculate than a large number of deterministic SAL scores and of being somewhat more robust with respect to changes in thresholds or observation uncertainties. Its main disadvantage is that it does not provide information on the ensemble spread.

OBS UNC-4: Considering observation uncertainty for verification of forecast ensembles

AUTHORS: Dorninger, Manfred; Kloiber, Simon.

Presenting author e-mail: manfred.dorninger@univie.ac.at; Affiliation (all authors): University of Vienna, Wien, Austria

ABSTRACT: Traditionally, verification means to verify a forecast (ensemble) with the truth represented by observations. The observation errors are quite often neglected arguing that they are small when compared to the forecast error. In this study as part of the MesoVICT (Mesoscale Verification Inter-comparison over Complex Terrain) project it will be shown, that observation errors have to be taken into account for verification purposes. The observation uncertainty is estimated from the VERA (Vienna Enhanced Resolution Analysis) and represented via two analysis ensembles which are compared to the forecast ensemble. For the whole study results from COSMO-LEPS provided by Arpae-SIMC Emilia-Romagna are used as forecast ensemble. The time period covers the MesoVICT core case from 20-22 June 2007. In a first step, all ensembles are investigated concerning their distribution. Several tests have been executed (Kolmogorov-Smirnov-Test, Finkelstein-Schafer Test, Chi-Square Test etc.) showing no exact mathematical distribution. So the main focus is on non-parametric statistics (e.g. Kernel density estimation, Boxplots etc.) and also the deviation between "forced" normal distributed data and the kernel density estimations. In a next step the observational deviations due to the analysis ensembles are analysed. In a first approach scores are multiple times calculated with every single ensemble member from the analysis ensemble regarded as "true" observation. The results are presented as boxplots for the different scores and parameters. Additionally, the bootstrapping method is also applied to the ensembles. These possible approaches to incorporating observational uncertainty into the computation of statistics will be discussed in the talk.

OBS UNC-5: Impact of gridded "truth" resolution on spatial verification scores

AUTHORS: Mittermaier, Marion.

Presenting author e-mail: marion.mittermaier@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: The Fractions Skill Score (FSS) has been used to routinely verify UK precipitation forecasts since early 2008. The primary focus has been on 6h precipitation totals. Since then the resolution of the regional model over the UK has changed from 12 to 4 to 1.5 km. During this period there was also a change from using a 5 km radar derived Quantitative Precipitation Estimate (QPE) to a 1 km version. These two radar products were available in parallel for a period of time when the 4 km model (UK4) was the operational model. This presentation explores the impact of the gridded data resolution on spatial verification metrics such as the FSS, and how the differences relate to whether detail is resolved in the model, in the „truth" data, both, or neither. In this study two neighbourhood sizes common to both truth resolutions are explored: 5 and 25 km. Note that the model is interpolated to the QPE grid. Percentile thresholds are used throughout, to remove the bias. Therefore the primary focus of the study was on the structure and spatial distribution of the precipitation, not the intensity. At 5 km scales the UK4 forecasts scored ~6% higher against the 1 km resolution QPE. This difference can be attributed to observation uncertainty due to interpolation and representativeness. The impact is greater for small neighbourhoods, reducing to around 2% for the 25 km neighbourhood.

CLIM-1 KEY NOTE: Drift in Decadal Predictions - A particular challenge for Verification

AUTHORS: Rust, Henning.

Presenting author e-mail: henning.rust@met.fu-berlin.de; Affiliation (all authors): Institut fuer Meteorologie, Freie Universitaet Berlin, Berlin, Germany

ABSTRACT: A climate projection is a climate model simulation with boundary conditions prescribed in terms of radiative forcing but starting from an arbitrary climate state on the model‘s attractor. Decadal climate predictions use the same boundary conditions but the climate model is initialised with a climate state close to the observed state. As this is not necessarily a state on the climate model‘s attractor, simulations show an artificial behaviour thought to be the model‘s attempt to get back to it‘s preferred states. This is artificial in the sense that it is neither a reaction to changing boundary conditions nor part of natural climate variability. From a post-processing perspective, this behaviour is conceived as a model bias changing with forecast lead-time. The term coined for this behaviour is “drift“. Typically, the drift is stronger when a full-field is used to initialise the model with a climate state compared to using anomalies only. However, drift can be present in both cases. Drift characteristics can furthermore depend on the climate state used for initialisation. Starting from the current recommendation for verifying decadal prediction (Goddard et al., 2013) and accounting for a drift (ICPO, 2011), we discuss alternative approaches to account for a drift. In particular, a parametric model for a drift changing with initialisation time is presented (Kruschke et al., 2015). This model is used to reduce drift in the ensemble mean and leads to an increase in skill. Within a forecast re-calibration framework, the underlying idea of a parametric model is then straightforwardly extended to additionally account for a conditional bias of the simulations and to adjust the ensemble spread to be representative for forecast uncertainty (Pasternack et al., in preparation). This approach allows to analyse and compensate for these various types of systematic errors and provides thus a valuable tool for model evaluation and forecast post-processing.

CLIM-2: Freva - Freie Univ Evaluation System Framework for Scientific Infrastructures in Earth System Modeling

AUTHORS: Kadow, Christopher; Illing, Sebastian; Schartner, Thomas; Kirchner, Ingo; Rust, Henning; Ulbrich, Uwe.

Presenting author e-mail: christopher.kadow@met.fu-berlin.de; Affiliation (all authors): Freie Universität Berlin, Berlin, Germany

ABSTRACT: The Freie Univ Evaluation System Framework (Freva - freva.met.fu-berlin.de) is a software infrastructure for standardized data and verification methods in Earth system science. Freva runs on high performance computers to handle customizable evaluation systems of research projects, institutes or universities. It combines different software technologies into one common hybrid infrastructure, including all features present in the shell and web environment. The database interface satisfies the international standards provided by the Earth System Grid Federation (ESGF). Freva indexes different data projects into one common search environment by storing the meta data information of the self-describing model, reanalysis and observational data sets in a database. This implemented meta data system with its advanced bu t easy-to-handle search tool supports users, developers and their plugins to retrieve the required information. A generic application programming interface (API) allows scientific developers to connect their verification tools with the evaluation system independently of the programming language used. Users of the verification techniques benefit from the common interface of the evaluation system without any need to understand the different scripting languages. Facilitation of the provision and usage of tools and climate data automatically increases the number of scientists working with the data sets and identifying discrepancies. Plugins are able to integrate their e.g. post-processed results into the database of the user. This allows e.g. post-processing plugins to feed statistical analysis plugins, which fosters an active exchange between plugin developers of a research project. Additionally, the history and configuration sub-system stores every analysis performed with the evaluati on system in a database. Configurations and results of the tools can be shared among scientists via shell or web system. Therefore, plugged-in tools benefit from transparency and reproducibility. Furthermore, if configurations match while starting an evaluation plugin, the system suggests to use results already produced by other users saving CPU/h, I/O, disk space and time. The efficient interaction between different technologies improves the Earth system modeling science framed by Freva.

CLIM-3: Joint uncertainty assessment of models and observations in verification of climate predictions.

AUTHORS: Bellprat, Omar; Massonnet, François; Siegert, Stefan; Menegoz, Martin; Guemas, Virginie; Doblas-Reyes, Francisco.

Presenting author e-mail: omar.bellprat@bsc.es; Affiliation (all authors): Earth Sciences, Barcelona Supercomputing Centre (BSC), Barcelona, Spain

ABSTRACT: Climate model simulations and observational references of the Earth?s climate are both uncertain representations of reality. While uncertainties in climate models have been assessed thoroughly by the modelling community, imperfections of observational references remain weakly explored in the forecast quality assessment of climate predictions. Practice of choosing an observational reference is rarely guided by objective criteria and relies often on data accessibility and institutional proximity. Uncertainty estimates provided by the products are further often ignored, partly because of the lack of formal concepts on how to consider observational uncertainties in model verification. This study illustrates implications of considering observational references as uncertain sources in the context of ensemble mean correlation skill of seasonal sea-surface temperature predictions. We will demonstrate that adopting the notion of uncertain observations allows to reverse the verification question: which observations have the smallest uncertainties given a suite of multiple model systems. The observational uncertainty originating from different observational references is consequently compared to sample uncertainties and shown to dominate these over many regions over the globe. Finally, we will show that accounting for dependencies between different models or observations is crucial to detect improvements in model systems or observational references. Massonnet, F., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J. (2016). Using climate models to estimate the quality of global observational data sets. Science. Bellprat, O., Massonnet, F., Siegert, S., Prodhomme, C., Macias-Gomez, M., Guemas, V. and Doblas-Reyes, F. J. (2017) Exploring observational uncertainty in verification of climate model predictions. Remote Sensing of the Environment, under review. Siegert, S., Bellprat, O., Ménégoz, M., Stephenson, D. B. and Doblas-Reyes, F. J. (2016). Detecting improvements in forecast correlation skill: Statistical testing and power analysis. Monthly Weather Review, (2017).

CLIM-4: Useful evaluations of variability and change in simulated climate for impact studies

AUTHORS: Ammann, Caspar; Brown, Barbara; Kalb, Christina; Bullock, Randy; Gilleland, Eric.

Presenting author e-mail: ammann@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research, Boulder, CO, United States

ABSTRACT: Climate impact studies rely on climate data that reflect the real-world conditions as closely as possible. Therefore, the criteria of what represents “good“ and useful driving data for such research is directly related to the specific issues of the applications. Hence evaluation tools must be flexible enough to be tailored to such needs. Traditional grid point-based comparisons between observational analyses and models, or between current and future climate, often do not reveal important information about the models? performance. Such summary statistics such as the correlation coefficient or the mean-squared error provide minimal information to developers, users, and decision makers regarding what is “right“ and “wrong“ with a model. For example, spatial or temporal displacements can be a reason behind a poor score despite ov erall reasonably faithful representation of the system; or, a good agreement in the mean doesn?t imply appropriate representation in variability or extremes. New spatial and temporal-spatial tools from the field of weather forecast verification have been adapted to more thoroughly answer some of the important earth system model evaluation questions. Examples of several applications of these tools in a climate context will be presented, such as spatial and temporal displacements in projections of El Nino-related precipitation and/or temperature anomalies using output of the NCAR-CESM large ensemble. In general, these tools provide diagnostic information about model performance accounting for spatial, temporal, and intensity differences that cannot be achieved using traditional (scalar) model comparison approaches. Thus, they can provide more information that can be used in decision-making and planning. Additionally, future extensions and applications of these tools in a climate context will be described.

S2S-1 KEY NOTE: Sub-seasonal to seasonal forecast Verification

AUTHORS: Vitart, Frederic; Ferranti, Laura.

Presenting author e-mail: Frederic.Vitart@ecmwf.int; Affiliation (all authors): ECMWF, Reading, United Kingdom

ABSTRACT: There is a growing interest in the scientific, operational and applications communities in developing sub-seasonal to seasonal forecasts (from weeks to a season) that fill the gap between medium-range weather and seasonal forecasts. Assessing their skill, and uncertainty, and exploring ways to communicate their benefits to decision makers are significant challenges. Sub-seasonal forecasting is at a relatively early stage of development, yet operational models are beginning to exhibit some skill based on a number of sources of sub-seasonal predictability (e.g. sea-ice, soil moisture, MJO, stratosphere-troposphere interactions, teleconnections, etc.). A major challenge in verifying sub-seasonal to seasonal forecasts is to take into account the flow dependency of the forecast skill and identifying forecast windows of opportunity. A main difference with short and medium-range forecasting is that the model systematic biases are too large to be ignored and re-forecasts are needed to bias correct the models. As the time range increases, the verification needs to be performed over increasingly larger spatial and time averages. The WWRP/WCRP sub-seasonal to seasonal prediction (S2S) database, which opened recently and which contains real-time S2S forecasts (available with a 3-weeks delay) and re-forecasts from 11 operational centres, represents an important resource to verify and compare operational S2S systems. Examples of sub-seasonal verification of the Madden Julian Oscillation, tropical cyclones and weather regimes using the S2S database will be presented.

S2S-2: Evaluation and Quality Control for the Copernicus Seasonal Forecast Systems

AUTHORS: Bhend, Jonas; Doblas-Reyes, Francisco J..

Presenting author e-mail: jonas.bhend@meteoswiss.ch; Affiliation (all authors): Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

ABSTRACT: The Copernicus activity QA4Seas (Quality Assurance for Multi-model Seasonal Forecast Products) aims at developing a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S). The quality assessment is user driven and will provide a comprehensive overview of the multi-faceted quality aspects of state-of-the-art seasonal forecast systems. User needs are collected through a survey with a broad range of stakeholders. These insights, together with general information about the data available, a schema to propagate all the relevant metadata to ensure full provenance of the products and the concepts that form the basis of seasonal forecasting are then used to specify the requirements for the Copernicus Climate Data Store (CDS). To illustrate the strategy proposed for the future development of the C3S, a prototype of the EQC system will be developed as part of the project. In this presentation, QA4Seas is introduced and first results of the evaluation of the multi-model seasonal forecasting system are presented. QA4Seas is promoted by the Barcelona Supercomputing Center (BSC-CNS), which leads a consortium of six institutions well known for their expertise in seasonal prediction, verification, and the provision of climate services.

S2S-3: Evaluating Sub-seasonal hindcast using an objective regionalization approach

AUTHORS: Recalde, Gloria; Zaitchik, Benjamin; Badr, Hamada.

Presenting author e-mail: grecald1@jhu.edu; Affiliation (all authors): Department of Earth and Planetary Science, Johns Hopkins University, Baltimore, United States

ABSTRACT: Subseasonal to seasonal (S2S) forecasts are an area of growing interest in both atmospheric research and applied climate services. Evaluation of dynamically-based S2S systems, however, is challenging. One significant shortcoming of standard evaluation methods is that they fail to adjust for potential spatial biases in model predictions: by evaluating models against observations at a common point, grid cell, or area, standard evaluation metrics severely penalize models that might capture the fundamental dynamics that lead to variability but happen to have a spatial bias in where precipitation occurs. Here, we present on an effort to address this limitation in application to Northwest South America (NWSA). S2S forecasts have the potential to contribute to preparation for both flooding and drought disasters in this region. We evaluate th e skill of selected NMME global forecast systems in NWSA through an approach designed to address spatial bias. The key feature of our approach is that prior to comparing models to observation we perform an objective climate regionalization for both the observed dataset and model forecasts. This allows us to identify analogous regions in observation and model that likely share common drivers of variability even if they are not perfectly collocated. Doing this provides a basis for evaluating models through a regional matching approach, in which model forecasts are compared to observations in the analogous region. In this pilot study, we perform regionalization using the Hierarchical Climate Regionalization (HiClimR) package for R. Regionalization is applied to a 30 year archive of NMME precipitation hindcasts and, as the observation dataset, the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset version 2.0. Model performance is then evaluated using both stand ard spatially matched methods and regionalization-based analogous region matching. Results provide insight on model behavior and provide a basis for adjusting for spatial bias when applying dynamically-based S2S forecasts.

7. Poster Presentations

P HIW-1: Extreme Verification ... relatively speaking

AUTHORS: Sharpe, Michael; Bysouth, Clare.

Presenting author e-mail: michael.sharpe@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: Growing interest in high impact weather has inevitably led to a desire to examine the ability of weather prediction models to forecast extremes; however, such an examination is particularly difficult because, by their very nature, extreme events are rare events. This presentation outlines recent work to develop methodologies for measuring how well Met Office probabilistic site-specific forecasts (a blend of ensemble models, including the ECMWF medium range ensemble) predict extremely hot summer days, extremely cold winter nights, extremely wet days and extreme wind events at UK synoptic sites. An analysis of the climatology at each station reveals that choosing a single threshold is inappropriate because it may almost never be exceeded at some locations but regularly exceeded at others. Therefore, in an attempt to forge a link between an event and its impact, site-specific event thresholds are chosen for each site by selecting the same percentile (corresponding to an event that occurs four times per year) from each climatological cumulative distribution function. Performance is evaluated using the Symmetric Extremal Dependency Index and two versions of the continuous ranked probability skill score (threshold weighted and partitioned).

P HIW-2: Verification of operational tropical cyclone forecast from 2010 to 2016 in western North Pacific

AUTHORS: Chen, Guomin.

Presenting author e-mail: chengm@mail.typhoon.gov.cn; Affiliation (all authors): Shanghai Typhoon Institute, Shanghai, China

ABSTRACT: Forecasts of tropical cyclone (TC) track and intensity from six official agencies and twelve operational models (seven global models and five regional models) during 2010 and 2016 are assessed to study current capability of track forecast guidance in the western North Pacific. The evaluation is performed from several aspects, including the mean, median, percentile distribution, regional distribution, relative position, homogeneous comparison, correlation analysis, and so on. Results show that the best models? performance already exceeded the mean level of subjective and approaching to the best subjective method. Noticeable improvement has been making year after year for a number of models, especially for NCEP-GFS. The performance of NCEP-GFS was almost close to ECMWF-IFS at the leading time levels from 24-h to 72-h. The regional distribution of position er rors shows that high-latitude region, low-latitude region (TC genesis region) and South China Sea are three main areas which large errors concentrated. A majority of NWP models are of their respective characteristics of systematic error through relative position analysis. Significant positive correlation exists between the initial forecast track errors and TC Pmin (TC size) for both ECMWF-IFS and JMA-GSM at the leading time levels before 48-h (around 30-h). Such a fact exhibits the heavy reliance of forecasts on the initial characteristic of TC.

P HIW-3: ECMWF and NCEP ensemble predictions verification of typhoon intensity during 2011-2015

AUTHORS: Wang, Haiping.

Presenting author e-mail: wanghp@cma.gov.cn; Affiliation (all authors): National meteorological center of China meteorological administration, Beijing, China

ABSTRACT: The ECMWF and NCEP ensemble predictions of typhoon intensity during 2011-2015 are explored. The ECMWF predictions have not improved in recent 5 years, which have the maximum absolute error values in 2015 because of the weakest intensity predictions. The error of NCEP average typhoon intensity prediction reduced significantly between 2013 and 2015, and which is below the average prediction error of ECMWF. The error increases with the increase of forecast time length. And ECMWF have more large error samples than NCEP. The collection of NCEP forecast error variations are not as regular as ECMWF?s, in other words, a lot of stronger and weaker forecast samples offset each other. In 2015 NCEP collection of the predictions in the early stage of the typhoon season forecast are weaker than observations, while stronger in the later. Overall average forecast was weak, while weak level is far less than the ECMWF. The stability of NCEP is also slightly better than ECMWF. The NCEP and ECMWF ensemble typhoon intensity forecast products during the rapidly intensification of typhoon are much weaker. In addition, although ECMWF forecast average error is bigger, but the number of better forecast samples is more than NCEP.

P HIW-4: Towards a user-oriented improvement of weather warnings. Opportunities and constraints for future developments

AUTHORS: Kox, Thomas; Lüder, Catharina; Kempf, Harald; Göber, Martin; Gerhold, Lars.

Presenting author e-mail: thomas.kox@fu-berlin.de; Affiliation (all authors): Freie Universität Berlin, Berlin, Germany

ABSTRACT: Based on five workshops with members of the German weather service and with forecast end-users from emergency management, water management, road maintenance, and the media, we discuss operational practices regarding weather information and specific end-user needs, along with future developments of the current warning system. The workshops were followed by a subsequent quantitative questionnaire survey (n5) with questions about current usage of weather information, (e.g. distribution, confidence, decision making) and open-ended questions about the implementability of new warning tools. Based on the results we derive opportunities and constraints for future developments. The work is closely related to the WEXICOM project funded by the Hans-Ertel-Centre for Weather Research, a research network of universities, research institutes and Deutscher Wetterdienst funded by the BMVI (German Federal Ministry of Transport and Digital Infrastructures).

P HIW-5: Verification of extreme weather warnings: How to define verifying events?

AUTHORS: Zingerle, Christoph; Hölzl, Simon.

Presenting author e-mail: christoph.zingerle@zamg.ac.at; Affiliation (all authors): Zentralanstalt für Meteorol gie und Geodynamik, Innsbruck, Austria

ABSTRACT: Issuing warnings of extreme weather is one of the fundamental and primary duties of a national weather service. First of all weather warnings are always predefined by meteorological conditions. As a consequence, such warnings not only trigger internal procedures inside the weather service, but also entail a number of actions in other national and local governmental services like civil protection agencies, hydrological services and any unit dealing with security or support for the general public. Furthermore, when it comes to warnings of extreme weather, in many countries national weather services are the single official voice. Hence, publishing such warnings is the most recognized activity of a weather service by media and warnings are frequently discussed in general public. In the end issuing weather warnings is a complex process, det ermined not only by the synoptic situation, but also by detailed internal specifications defined in some kind of warning manual (definition of warning levels by climatology, return levels, operational handling, ?) and the differing requirements of external users such as civil protection or hydrological services. Knowing the overall performance of a warning system is of importance when it comes to the discussion of quality of a service in responsibility to public authorities. On the other hand, handling of individual warning cases is crucial for internal discussions and debriefing with external users after a warning season. When actual extreme weather events are defined for operational verification of weather warnings, the complex process ending up in issuing a warning by forecasters needs to be reproduced based on the meteorological observations in some way. In our operational setting, the verification algorithm matches issued warnings with so-called events that are defined in an initial step. In the case of heavy precipitation, rainfall events are separated by dry events (or "null events"). Definition of storm events is carried out in a similar way, whereas periods of maximum wind gusts above and below 60 km/h are separated. The minimum time period of dry or non-storm events separating rain fall or storm events, respectively, is set to 6 hours. Warning levels (defined by climatology) are assigned to the respective events, depending on rain amount per time period (individual return levels per accumulation time for every community) or maximum wind gusts (return level for single event for every community). When matching warnings with events, margins of duration and amplitude of every single event allow to introduce some kind of fuzziness to the verification and reflect the complexity of the warning process. Further adjustments are done, if a definitive match is not possible. Finally, contingency tables are constructed and skill scores are calculated for individual communities and entire districts. The operational verification of manually issued warnings for heavy rainfall and storm events is available for the years 2012 to 2016, when the current warning system has been used in operations. Warnings are verified against events defined by the algorithm from measurements from about 250 automatic weather stations operated by ZAMG. Verification was carried out on community basis, where communities situated within a 5-kilometer-radius from a weather station were taken into account. Warnings with a minimum lead time of 12 hours were considered for the verification process. In a next step, the verification routine for heavy rainfall events will be extended to data from the INCA (Integrated Nowcasting through Comprehensive Analysis) System, providing a precipitation analysis for Austria on a 1-by-1-kilometer grid based on station and radar measurements every 15 min.

P USR VAL-1: Space - the final verification frontier?

AUTHORS: Sharpe, Michael; Bingham, Suzy; Jackson, David.

Presenting author e-mail: michael.sharpe@metoffice.gov.uk; Affiliation (all authors): Met Office, Exeter, UK

ABSTRACT: Changes in solar activity can lead to significant impacts on earth, including power outages caused by geo-magnetically induced currents and radio communication loss due to ionospheric disturbances. Therefore, the Met Office Space Weather Operations Centre was established to assist affected industries and infrastructure in the UK build resilience to space weather. One of the centre‘s daily products is a four day probabilistic forecast of X-ray flares, high energy electrons, geo-magnetic storms and high energy protons. It is crucial for forecasters, forecast-users, modellers and stakeholders to understand the strengths and weaknesses of these forecasts and this presentation outlines the methodologies used to verify the X-ray flare and Geo-Magnetic Storm components. To assess forecast skill it is helpful to compare against a reference; various rolling prediction periods have been considered for this purpose and the time-period which minimises the RPS is chosen as the reference. Analysis of rolling 12-month performance suggests that events tend to be over-forecast and only the first day provides a better prediction than the reference.

P USR VAL-2: Improved Airport Forecast Verification for Forecasters

AUTHORS: Jelbart, Joe; Butcher, Paige.

Presenting author e-mail: j.jelbart@bom.gov.au; Affiliation (all authors): Bureau of Meteorology, Melbourne, Australia

ABSTRACT: A standard coded TAF (aerodrome or airport forecast) potentially contains a lot of information about a wide range of environmental conditions, with weather and associated phenomena being one of many elements. If a corresponding set of observations can be obtained, then a very comprehensive verification system can be developed. The challenge presented by the AVS2 (Aviation Verification System 2) project was to deliver an operational verification and business intelligence service that provided user-oriented performance metrics and business intelligence to forecasters, managers and industry stakeholders. At the same time, it was necessary to investigate a fresh approach to TAF verification, so a simultaneous process of R&D resulted in new algorithms to support a new core methodology. The AVS2 algorithm architecture has been developed with the ability to incorporate existing and future business requirements. The algorithm architecture of AVS2 is made up of many algorithms chained together as a data transformation pipeline. Initial data extraction is followed by a preprocessing and cleansing stage; next, the TAF and all available observational data are used to build a single unified data model, which standardises and links data based around the concept of a TAF hour. This data model makes all downstream algorithms simpler by normalising the categorisation and structure of data, but not the data values. The unified data model is used to build an object model representing the relationships between all TAFs and observations. The object model is in turn used to populate a Verification Metrics Unit Table (VMUT) - a data object that is a universal representation of a TAF hour and can be used for statistical object arithmetic, in a similar way to date/time objects in most programming languages. There are three types of VMUT. The basic VMUT is a data object that records discrete comparisons, i.e., discrete observations against discrete forecasts. These VMUTs are each converted into a Dichotomous VMUT according to a tunable algorithm - these represent how the TAF hour will be accounted for in a statistical analysis and can only contain zeros and ones. The final conversion adds one or more Dichotomous VMUTs together resulting in a single Performance VMUT. All performance metric formulae, both established and future, can be applied to a Performance VMUT, which simplifies each individual formula and clearly defines the data set and formula variables. The AVS2 algorithm architecture lays the foundation for a relatively easy way to get an identical report based on persistence or climatology for direct comparison with a standard TAF report. Developing the AVS2 algorithm architecture was an opportunity to explore how data models, data standardisation and conversion algorithms can be used to enhance and simplify the process of automated forecast verification. With the first phase of the AVS2 project complete, the Bureau of Meteorology has a quasi-operational system that routinely sends verification report emails (business intelligence) to forecasters and their managers, which can be used for decision support, rapid fine tuning of an individuals forecasting and improving service quality.

P USR VAL-3: Study on meteorological grade of the Forest Fire Danger and rationality verification

AUTHORS: Yuan, Meiying; Xu, Nanping; Zhou, Xiujie; Qi, Duo.

Presenting author e-mail: meiying_yy@126.com; Affiliation (all authors): Provincial Meteorological Observatory of Heilongjiang, Harbin, China

ABSTRACT: With an forestry area of 240,850,000 mu (1 mu=0.0667 hectares), Heilongjiang Province is a key province of forest fire prevention in the country. The research result revealed in this paper indicates that: 12 meteorological factors are closely related to fire danger ratings, among which 5 are current factors, 7 are previous factors and 2 are vegetation factors. Based on fire data of many years and their relationship with meteorological factors, we collected historical materials and real time observation materials of 73 stations across the province in two periods: spring prevention period (March-June) and autumn prevention period (September-November). After screening and calculation, we created indicators and mathematical models for 12 single-factor fire danger contribution degrees. Furthermore, through aggregative indicator method, contribution degrees for single-factor fire danger were summarized into aggregative indicators with practical application value. Then, 4-level discrimination was adopted: (1) Calculated 12 single-factor fire danger contribution degrees and built correspondent mathematical models for various meteor ological factors. With reference to mathematical model of fuzzy mathematics, we established contribution degree model of factors; (2) Calculated the contributions of current factors and previous factors to fire danger; (3) Determined fire danger ratings with multi-factor aggregative indicator method; (4) 4-level discrimination was critical for the study of forest fire danger weather ratings and was determined by the relationship between the frequency of forest fire occurrence and fire danger ratings. The standard for 5-level fire danger weather ratings was generated after 4-level discrimination. The rating of fire danger should abide by the following principles: (1) Level 1 (non-inflammable). The frequency of fire occurrence should be controlled within 5%. (2) Level 2 (difficult-inflammable). The frequency of fire occurrence should be controlled within 10% (generally less than 10%). (3) Level 4 and Level 5. The frequency of fire occurrence should be 50% or above. These are just basi s principles. In actual rating, province-wide uniformity should be taken into consideration. Based on the work above, corrections were conducted. The corrections successively included: snow correction, spring plant regreening stage correction, autumn withering correction and frost stage correction, special factor correction. Meteorological grade of the Forest fire danger established by the comprehensive index system above mainly adopts the following four methods. (1)the verification on frequency of different grade; (2) the verification on the probability of different grade; (3) accuracy test; (4) weight fuzzy assessment test. The results showed that the meteorological grade of forest fire danger accurately described the weather conditions of reality. The grade has wrote into China's Standard QX/T 77-2007. Key words: Forest fire danger; meteorological grade; comprehensive index; rationality verification;

P USR VAL-4: TREND Forecast Verification in the MET Alliance

AUTHORS: Mahringer, Guenter.

Presenting author e-mail: guenter.mahringer@austrocontrol.at; Affiliation (all authors): Austro Control MET Service, Hoersching, Austria

ABSTRACT: TREND forecasts are very short range forecasts used in aviation. They consist of a concise statement of the expected significant changes in the meteorological conditions (wind direction, wind speed, visibility, present weather and cloud ceiling) at an aerodrome. They are to be appended to a local routine or local special report, or a METAR or SPECI. Significant changes are defined by thresholds (e. g. for visibility, cloud ceiling height and wind speed) significant for aviation operations. TREND forecasts are usually issued every 30 minutes and the period of validity is 2 hours from the time of the respective report. TREND forecast verification focusses on the question if significant changes that occurred within the period of validity have been forecast, and if forecast changes in fact occurred. In 2016, six members of The MET Alliance have agreed on a TREND verification project. Based on a method developed in Austro Control, TREND forecasts for 37 airports in Ireland, Luxembourg, Germany, Switzerland, The Netherlands and Austria are verified routinely. The verification method is to compare best and worst forecast conditions within the period of validity with best and worst observed conditions for each class of initial conditions. Thus, contingency tables for forecasts / observations of significant changes are derived, for which standard verification scores are computed. The results for the first project year show that producing good TREND forecasts in an operational environment requires good Nowcasting tools, methods and trained forecasters. Furthermore, a convenient setup of the forecasting process is required by which the forecasters´ focus is concentrated on the task. It is shown that the deterioration of conditions is more in the forecasters´ focus than the improvement for some types of changes. For example, forecasters are much more focussed on forecasting the onset of a thunderstorm than its end. Verification results are used in quality management, forecaster training, in the development of forecasting methods and for improving operational environments.

P META-1: MAE and RMSE: which is the better measure of model performance?

AUTHORS: Brassington, Gary.

Presenting author e-mail: g.brassington@bom.gov.au; Affiliation (all authors): Bureau of Meteorology, Sydney, Australia

ABSTRACT: The mean absolute error (MAE) and root mean square error (RMSE) are common metrics applied to assess model performance. We note that in general, RMSE^2 = MAE^2 + VAR such that the RMSE includes both the MAE and the variance (a biased estimator) of the absolute errors. The sensitivity of RMSE to a small number of outliers can be attributed to this variance. Additional statistical properties for both metrics are derived based on the assumption that the sample errors are normally distributed. For an unbiased (or bias corrected) model both MAE and RMSE are shown to estimate the total error standard deviation to within a constant coefficient such that MAE ? ?(2/?) RMSE . Both metrics have comparable behaviour in response to model bias and asymptote toward this bias as the bias increases. MAE (RMSE) is shown to be an unbiased (biased) estimator. The total error standard deviation, TESD = ? (?/2) MAE ± 2/?k ?(?/2-1) ?(?/2) MAE, is recommended for forecast verification, with a 95% confidence interval defined as a scaling of the MAE itself. A sample size (k) on the order of 400 and 40000 provides an approximate error scaling of 10% and 1% respectively. For delayed mode model verification where quality-controlled observations are available, the RMSE sensitivity to large model errors is advantageous. An analogous estimate of TESD based on RMSE is derived and compared. For comparisons of models, it is also instructive to compute both the MAE and RMSE and present this in a model performance diagram.

P META-2: The verification of typhoon forecast results by global model

AUTHORS: Kim, Ki-Byung; Lee, Eunjeong; Lee, Eun-Hee.

Presenting author e-mail: kb.kim@kiaps.org; Affiliation (all authors): Korea Institute of Atmospheric Prediction Systems, Seoul/Dongjak-gu, Korea

ABSTRACT: In Korea Institute of Atmospheric Prediction Systems (KIAPS) develop the global model and operates in semi-real time for forecast. Verification of that is essential to analyze and improve the model. In this study, we focused on examine track and intensity of the typhoon that affect the human lives with strong wind and heavy rainfall, and verify the predicted results of that genesis over the western north Pacific in 2016. The typhoon center is tracked by Geophysical Fluid Dynamics Laboratory (GFDL) vortex tracker, it is a program that searching for the average of the maximum or minimum of several parameters in the vicinity of an input first guess position of the targeted vortex. After the using that program, we handle the output data and verify against with Joint Typhoon Warning Center (JTWC) best track data. We aimed at a typhoon that activated for m ore than three days, and looked at the simulation performance according to the predicted start time.

P META-3: CRA Vs MODE for Heavy Rain Cases over India

AUTHORS: Ashrit, Raghavendra; Sharma, Kuldeep; Singh, Harvir; Mohandas, Saji.

Presenting author e-mail: raghu.ashrit@nic.in; Affiliation (all authors): National Centre for Medium Range Weather Forecasting, Noida (U.P), India

ABSTRACT: Rainfall is probably the most important and challenging of all the NWP products. Rainfall forecasts are expressed in terms of the amount of rainl accumulated over a specified time period which is known as quantitative Rainfall forecast (QPF). Though QPF has improved significantly over the last decades, it still remains the most difficult problem in operational NWP Model forecasts with large uncertainties. While the large scale (sub-continent) and long period average (seasonal) rainfall amounts are reasonably well predicted, significant forecast errors are evident over smaller spatial (basin or sub-basin scale) and temporal (medium range) scales. Often models fail to correctly predict the rainfall amounts, or spatial location or both. Statistical quantification of the forecast errors is often carried out using standard verification metrics using categorical scores. Fu rther, for spatial verification various advanced methods have been proposed in recent years. Contagious Rain Areas (CRA) and Method for Object-Based Diagnostic Evaluation (MODE) are object-based methods which seek to identify matching features in observations-forecast pairs and quantify the forecast errors. This paper provides an intercomparison of the two methods and summarizes the differences and similarities with application of these methods for three cases of heavy rain events during 2015 and 2016. It is found that both MODE and CRA have advantages and demerits within their domain and what they want to achieve. Advantage of CRA over MODE is that CRA method allows for decomposition of the forecast error in terms of contribution from error in pattern, volume and location which is not available in MODE. Advantage of MODE over CRA is that MODE also computes angle difference which is missing in the CRA method. Both methods together provide an extended set of spatial verification metrics (in addition to traditional scores) that could be useful for monitoring the model performance and its improvements in rainfall forecasts.

P OPS SOFT-1: The overview of the verification systems of the KIAPS next generation model

AUTHORS: Kwon, Young; Lee, Eun-Hee.

Presenting author e-mail: yc.kwon@kiaps.org; Affiliation (all authors): Korea Institute of Atmospheric Prediction Systems, Seoul, South Korea

ABSTRACT: The next generation global model has been under developed by Korea Institute of Atmospheric Prediction Systems (KIAPS) since year 2011 (KIM, KIAPS Integrated Model, Choi and Hong, 2016, Song et al, 2017, Park et al, 2016 and so on). As the development progressing, the systematic verifications become important not only for further model improvement but also for measure of maturity of the model. KIAPS takes two approaches to verify the performance of the developing next generation model. One is objective verification using usual metrics such as root mean square error, bias error, threat score, and other statistical measures in order to quantify the strength of weakness of the model. Also, simulations with some standardized cases "testbed cases" are conducted when the model is updated to ensure the expecting improvement. The other is subjective verification method by which the model performances are evaluate from the point view of forecasters. The subjective verification compliments the weakness of the objective verification aforementioned and this method assure the ultimate usefulness of the model which is providing forecast guidance to the operational forecasters. In this talk, the brief summary of the overview of the whole KIAPS verification system with some results will be presented.

P OPS SOFT-2: The update of NMC operational verification system and results analysis in the past 26 years for upper-air products of deterministic NWP

AUTHORS: Li, Li.

Presenting author e-mail: 1249205067@qq.com; Affiliation (all authors): National Meteorological Center, beijing, China

ABSTRACT: The Commission for Basic Systems (CBS) Coordination Group on Forecast Verification (CGFV) finished the new verification standards and the operational production of new scores in the new file format for upper-air forecasts of deterministic NWP products produced by GDPFS centers in 2012. So the National Meteorology Center (NMC) in China has updated the operational verification methods and programs for the upper-air forecasts of the deterministic NWP products according to the CBS new verification standards. Firstly, this paper detailed the new CBS verification standards and procedures. Secondly, the differences between statistical scores got from old verification system and new verification system would be compared in detail, including the influences produced by the change of the climatology and verification resolution. Lastly, through more than twenty years statistics results, the forecast level and skill of upper-air elements of the deterministic NWP from T42L9 model in 1990 to GRAPES (Global/Regional Assimilation and PrEdiction System) developed independently by China NMC in 2016 would be analyzed and summarized systematically.

P OPS SOFT-3: Using Python tools to deliver flexible, repeatable verification to support evidence driven forecast automation

AUTHORS: Ioannou, Ioanna; Liu, Maoyuan; Jack, Harry; Griffiths, Deryn; Foley, Michael.

Presenting author e-mail: ioanna.ioannou@bom.gov.au; Affiliation (all authors): Bureau of Meteorology, Melbourne, Australia

ABSTRACT: The Australian Bureau of Meteorology is working to partially automate forecasts to allow for more efficient use of forecaster time. Achieving this objective requires careful verification to ensure we understand the changes that automation would bring to the quality of forecasting services. Our team is producing the necessary verification to identify when and where automated forecasts are of sufficient quality and when improvements are necessary. To carry out this verification it was necessary to produce a broad range of verification metrics with many spatial and temporal stratifications and communicate this information to the relevant stakeholders. This process needed to be easily repeatable for different versions of automated forecasts. To satisfy these requirements our team developed a Python framework that supports easy and flexible: a) retrieval of forecast and observation data, b) calculation of a broad range of standard and novel verification metrics, c) saving and querying of verification results, and d) visualisation and communication of results to stakeholders through web interfaces. The framework is built upon standard open source Python tools that are easily accessible to researchers. For example, Jupyter Notebooks serve as an environment for fast analysis and prototyping, the xarray library is used to easily match forecasts to observations and compute verification metrics, while Jupyter Dashboards provide interactive web-based exploration of verification results. The aim of this presentation will be to demonstrate how we have harnessed such Python tools to make verification easy, scalable and reproducible. We will provide a walkthrough of the tools that support each step of our verification process and show other researchers how they can start using them in their own verification work.

P OPS SOFT-4: s2dverification: an R package for climate forecast verification

AUTHORS: Manubens, Nicolau.

Presenting author e-mail: nicolau.manubens@bsc.es; Affiliation (all authors): Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), Barcelona, Spain

ABSTRACT: s2dverification v2.8.0 is an open-source R package for the quality assessment of seasonal to decadal forecasts using state-of-the-art verification scores. The package provides tools for each step of the forecast verification process: data retrieval, processing, calculation of verification measures and visualisation of the results. Examples are provided and explained for each of these stages using seasonal output from the EUROSIP multi-model set (GloSea-5, ECMWF System4, Météo-France System4, NCEP) over Europe.

P OPS SOFT-5: The key programme of CMA for Development of Forecast Verification Technique (CMA-FVRT), an introducton

AUTHORS: Chen, Jing; Xue, Feng; Dai, Jianhua; Li, Li; Wei, Qing; Zhao, Bin.

Presenting author e-mail: chenj@cma.gov.cn; Affiliation (all authors): Numerical Weather Prediction Center, CMA, Beijing, CHINA

ABSTRACT: A key programme, The Development of Forecast Verification Technique(CMA-FVRT), was approved by Chinese Meteorological Administration (CMA) in the end of 2016. The purpose of this programme is to enhance the capabilities of verificaton in the field of new operational forecasts which have been established by CMA in recent years. This key programme(CMA-FVRT), which was lauched in Jan. 2017 and will finish in 2020, has been organized by National Meteorological Center of CMA as well as six provincial weather office. The main tasks are: (1) Standardizing the observations for verification purpose based on CMA CIMISS data base. (2) Expanding the new verification methods to early warning and nowcasting forecast products. (3) Developing the method of convective-permitting NWP forecast and refined grid forecast products. (4) Establising the operational verification system of for environment, marine and hydrological forecasts in CMA. (5) Improving the technique and operational verification system for typhoon forecast products. An introducton of this key programme is here provided. The aims and main tasks will be presented and future steps will be outlined.

P OPS SOFT-6: Adaptive population selection for verification data mining

AUTHORS: Brassington, Gary; Divakaran, Prasanth; Beckett, Duan.

Presenting author e-mail: g.brassington@bom.gov.au; Affiliation (all authors): Bureau of Meteorology, Sydney, Australia

ABSTRACT: We would like to introduce an innovative approach to generalise the access to verification data archives that enable the non-specialist to derive highly tailored sub-populations, and apply a set of standard statistical operators. The target of our verification is databases of global ocean forecasts projected onto the observation space. Observations are unstructured and more difficult to work with when compared with gridded products. Our approach is to provide a series of innovative web-based tools (based on open standards) to control the selection of sub-populations in both space and time in order to simplify the construction of verification statistics. This allows a user to create highly tailored polygons for their region of interest with supporting information of the sample size and distribution in space and time. We demonstrate our approach through an internet accessible website based on a virtual machine (http://130.56.244.252/monitoring/index.php?pg=class4). This website is underpinned by the verification database that is accessible via the internet using an OPeNDAP server. The power of this product is that it enables the non-specialist to general data mining of a large and growing database, and helps to move away from pre-computed summary statistics. Moreover, this framework also include provision for user to include more comprehensive statistics options such as confidence intervals, greater control over graphical exports, greater range of personal and group preferences, greater integration with the model product, and additional reference data types.

P OPS SOFT-7: Emet or forecast verification using the power of a relational database

AUTHORS: Lemay, Francois; Husson, Thierry.

Presenting author e-mail: francois.lemay@canada.ca; Affiliation (all authors): Environment-Canada and Climate Change, Montreal, Canada

ABSTRACT: Anyone who does forecast verification has to handle a lot of different types of data, all kinds of observation sources as well as different types of models, different not only by their caracteristics but also in there proufound nature (NWP, post-treatment, probabilistic, ensembles, etc.). And God knows how mind puzzling it might be! Especially when one wants to do types of verification that requires crossing different aspects of the forecasts (ex.: say multivariable verification or conditional verification). Developped by a small team at Environment-Canada and Climate Change (ECCC), the Emet system whose corner stone is a powerful postgreSQL relational database allies python driven SQL and R scripts to ingest forecast verification data into a database, cross them with relevant sources of observations, do requests on those data by periods, geographical domains, variables, and produce hundreds of different possible verification products for different scores and different representations of them. Users can then consult their forecast verification results using a friendly user application through a simple web browser. But the real power of Emet is the handling data capacity of the postgreSQL database that allows one to apply conditions and cross all different kinds of data wheither they're observations, analysis or forecasts. And therefore allows one not only to evaluate the forecast peformance but also to corner the specific weather situations of interest.

P OPS SOFT-8: Enhanced Verification Capability within the Developmental Testbed Center

AUTHORS: Jensen, Tara; Halley Gotway, John; Fowler, Tressa; Brown, Barbara; Bullock, Randy; Strong, Bonny.

Presenting author e-mail: jensen@ucar.edu; Affiliation (all authors): National Center for Atmospheric Research/Research Applications Lab, Boulder, Colorado, USA

ABSTRACT: Robust testing and evaluation of research innovations is a critical component of the Research-to-Operations (R2O) process and is performed for the U.S. National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Prediction (NCEP) by the Developmental Testbed Center (DTC). At the foundation of the DTC testing and evaluation (T&E) system is the Model Evaluation Tools (MET). MET is a state-of-the-science verification package supported to the community through the DTC. The verification team within the DTC has been working closely with DTC teams as well as the staff at NCEP to enhance MET to better support both internal T&E activities and external testing within the community. This presentation will demonstrate several advancements that were made available in the current release. These include ability to run m ultiple convolution radii and thresholds with one call to the Method for Object-based Diagnostic Evaluation (MODE), the enhancement of MODE to follow objects through time (MODE-TD), enhancements to facilitate storm- or feature-centric evaluations, the inclusion of cosine latitude and grid area weighting for larger domains, support of several cloud analysis fields and satellite-based cloud lidar fields, the addition of the High Resolution Analysis (HiRA) methodology for point observations, and python scripting to facilitate running MET systematically.

P OPS SOFT-9: How MeteoSwiss measure the overall quality of its forecasts

AUTHORS: Letestu, André-Charles; Cattani, Daniel.

Presenting author e-mail: andre-charles@meteoswiss.ch; Affiliation (all authors): MeteoSwiss, Geneva, Switzerland

ABSTRACT: Since 2008, MeteoSwiss performs a verification of the quality of its forecasts. In 2013, a new global score was designated for administrative purposes as a communication tool aiming to provide a broad panel of non-specialists with an overall measurement of the quality of general deterministic forecast provided by MeteoSwiss. In parallel, a verification system addressed to forecasters was developed. It allows the forecaster to compare his or her score to the models for various ranges, parameters and weather situations.

P SPATIAL-1: The ECMWF model precipitation systematic error in east of Southwest China based on the Contiguous Rain Area Method

AUTHORS: Jiaolan, Fu.

Presenting author e-mail: fujiaolan@cma.gov.cn; Affiliation (all authors): National Meteorological Center of China Meteorological Administration, Beijing, China

ABSTRACT: Contiguous Rain Area Method (CRA for short) is an object-oriented verification procedure. By setting a threshold for precipitation, the contiguous rain area is identified and isolated, and the forecast error of the rain area could be analyzed, which can avoid the "double penalty" of traditional score method. The horizontal displacement is determined by translating the forecast rain field to the observed. This allows a decomposition of total error into components due to: location, rain volume and pattern. ECMWF's forecast products are one of the best guidance for National meteorological center in China. However, there are still some forecast biases in the location, amount and pattern of rainfall, especially in Southwest China, which is characterized by torrential rain in warm season. Due to the complex feature of terrain and plenty of moisture and thermal energy, the rainfall in Southwest China is largely caused by mesoscale convective systems and influences of topography on orographic precipitation are general. According to the experience of local weather forecasters, ECMWF's rainfall products in the short and median range usually present a trend of displacement bias to the observed one. ECMWF's rainfall forecast still provides a good reference to the forecaster if they know the systematic error, so it is needed to find the real error of the forecast, which will help to build up an error base referred by the forecaster to calibrate or adjust the model forecast. Typical heavy rainfall events during May to September in 2011-2014 over east of Southwest China are chosen to figure out the displacement and intensity bias of the ECMWF's rainfall forecast. Initially, the overall errors for all cases are analyzed. Then, based on the classification of circulation and synoptic systems, all cases are divided into several types, the difference of the errors between them are also discussed. 119 CRAs are found by using the CRA technique. Their location and intensity errors of forecasted by ECMWF model in 36h are analyzed by the CRA verification method. Then these cases are classified into three types according to the weather systems, they are low vortex and shear line over east of Southwest China (the first type), shear line located over Jianghuai Basin and east of Southwest China (the second type), and southerly wind type (the third type), respectively. The systematical error of each type is presented. The results show that: The pattern error for all CRAs is most great, about 60% of total error, then is the location error, about 30%; the volume error is the least, about 10%. The average location error is 0.7º westward to observed, and the meridional displacement error is not clear. The mesoscale-? rain area may be under-forecasted by the model; however the horizontal scale larger than mesoscale-? rain events would be over-forecasted for its rain area and total rainfall volume and under-forecasted for the average rainfall intensity and maximum rain. The intensity error is similar for the first and the second type, in which the rain area is over-forecasted and the average rain intensity and maximum rainfall is under-forecasted. However, the intensity is under-forecasted or the event is totally missed for the third type. For the location error, there is westward for the first type, northwestward for second type, and southwestward for third type.

P SPATIAL-2: The contiguous rain area (CRA) method application for the MesoVICT cases in the framework of the COSMO INSPECT project

AUTHORS: Bundel, Anastasia; Muravyev, Anatoly.

Presenting author e-mail: a.bundel@gmail.com; Affiliation (all authors): Hydrometcentre of Russia, Moscow, The Russian Federation

ABSTRACT: The CRA method using the R SpatialVx package is applied for the MesoVICT cases. Different precipitation accumulations and thresholds are verified. The influence of the matching function choice is studied. The outcomes are compared with those obtained earlier for the Caucasus Sochi region, another mountainous terrain. Preliminary results for the method application to the ensemble data are given.

P SPATIAL-3: Evaluation of two convective scale version of the Unified Model using two spatial verifcation methods

AUTHORS: Phaduli, Elelwani.

Presenting author e-mail: elelwani.phaduli@weathersa.co.za; Affiliation (all authors): South African Weather Service, Pretoria, South Africa

ABSTRACT: The South African Weather Service (SAWS) is currently using the Unified Model (UM) as the main model for NWP purposes. The UM is run at SAWS under the license agreement between SAWS and the United Kingdom Meteorological Office (MO, hereafter). SAWS started using the UM with 12km horizontal grid spacing (SA12 hereafter) in 2006, SA12 has been run once a day at 00:00UTC over the Southern African domain (0-44S; 0-56E) with 48hr lead time and grid spacing of about 0.11 degrees. Currently with the commissioning of the high performance supercomputer, SAWS is able to run two high resolution convective scale or convective permitting models over Southern Africa one with 4km horizontal grid spacing (SA4) and the other with 1.5 km horizontal grid spacing (SA1p5). The SA4 configuration is run over the Southern African domain (0-38S; 5-54E) with 72hr lead time and grid spacing of about 0.038 degrees. The SA1p5 configuration is run over the South African domain (36-22S; 15-34E) with 36hr lead time and grid spacing of about 0.013 degrees. Both convective scale models are run four times daily at (00:00UTC, 06:00UTC, 12:00, 18:00UTC) with initial and boundary conditions derived from MO's Global Atmosphere model (GA6) forecasts. Numerous studies have shown that improving the horizontal resolution of an NWP model can bring more benefits to the skill of the model forecasts. Weismann et al (1997), used the non-hydrostatic cloud model to simulate the squall lines using 1-12 km resolution models, they found that the squall lines became more realistic as the horizontal grid spacing was decreased to 4km. A study by Davis et al 1999 showed that decreasing the grid spacing of the MM5 model to 1.5km led to the improved diurnal circulations produced by the topography and varying land surface conditions over West central Utah. Another study done at the UK Met Office using the high resolution configurations of the UM indicated that the 4km and the 1km grid spacing models often give more realistic looking precipitation fields because convection is not parameterized but is represented explicitly in those models. Learn et al. (2008). Though increasing horizontal resolution of NWP models can be more beneficial to the forecasters and other model users, improved verification strategies are still required especially for evaluating Quantitative Precipitation Forecasts. There are drawbacks associated with using continuous and categorical verification methods to evaluate high resolution NWP models. Continuous and categorical verification scores often fail to account for the unique characteristics of precipitation, which is attributed to the fact that continuous verification scores are sensitive to discontinuities and outliers and categorical verification scores are sensitive to bias and the base rate of the event. In this study two spatial verification methods are used to evaluate different convective scale configurations of the UM model. The first method used was the Method for Object-Based Diagnostic Evaluation (MODE) which is an object-based verification approach for comparing gridded observations with gridded forec asts. Davis et al., (2006). The other approach that will be used is the intensity scale method, as introduced by casati et al. (2003). Preliminary results has shown that the addition of two convective scale configurations of the Unified Model at the South African Weather Service has led to improvement in both model forecasts and downstream products. Verification statistics of the 1.5 km convective scale model (SA1p5) and 4km convective scale model (SA4) convective scale models has shown that the two models performs better than the operational 12km model(SA12).

P SPATIAL-4: Object-Based Verification and Evaluation for different types of Severe Convection Forecasting Products

AUTHORS: Mao, Mao; Dai, Jianhua; Li, Baiping; Zhang, Xin.

Presenting author e-mail: mao0104mao@sina.com; Affiliation (all authors): Shanghai Central Meteorological Observatory, Shanghai, China

ABSTRACT: According to the characteristics of severe convective weather and the requirements to focus on various factors, object-based verification method is developed to find the potential value of forecasts. First of all, convection objects, such as convective cells and severe convective weather area, are identified in both forecasts and observations based on intensity and area which meet certain requirements. Then objects in the two data-sets are matched according to the area, distance and morphology. After that, objects in forecasts are verified in five aspects, grade TS score, grade size, distance of center of gravity, cross-correlation and morphology (axial and ellipticity). Finally, based on the evaluation preference of users, verification scores by weighted average and evaluation for area, position and shape are provided. Three types of severe convection forecasting products, such as QPF, REF, or convection probability product of the Chinese Meteorological Administration (CMA) SWAN (Severe Weather Analysis and Nowcasting) system, are verified by this object-based method. This method can provide quantitative verification in coincidence and deviation for area, position or intensity. It also can explain why the regular verification scores (such as TS score) are low. More effective verification and evaluation information, combined with above elements, is provided for forecasters.

P SPATIAL-5: Spatial features of QPF obtained from an extension of dichotomous verification at grid point level

AUTHORS: Bravo Blanco, Manuel; Moré, Jordi; Mercader-Carbó, Jordi.

Presenting author e-mail: mbravob@meteo.cat; Affiliation (all authors): Meteorological Service of Catalonia (SMC), Barcelona, Spain

ABSTRACT: When verifying quantitative precipitation forecasts (QPF) in numerical models, it is a common practice to convert forecasts in dichotomous categories of yes/no events. Then, contingency tables are computed at different precipitation thresholds to derive verification indexes for a whole domain (POD, FAR, BIAS, and CSI among others). This simple methodology is very useful to provide global information, but it can miss important aspects of the forecast skill at particular regions within the domain. Thus, for instance, a global overestimation of precipitation at a certain threshold can hide areas where underestimation may be the general rule. For this purpose, the Meteorological Service of Catalonia (SMC) has recently introduced the computation of dichotomous contingency tables defined at every grid point along with the traditional ones which consider the overall grid. In the present study, the daily operational outputs of the WRF-ARW model are verified using this new approach and compared with the standard ones for the whole year 2016. The comparison reveals that QPF‘s performance varies according to the specific region of model domain and thus provides useful information for end users and model developers as well. This simple re-use of well-known verification indexes has the main advantage of keeping spatial information. Additionally, they are easy to understand (user oriented verification) and to implement. Moreover, this approach can be extended to other variables and even to fuzzy verification. The main limitation is that larger datasets are required compared to the traditional use of these indexes since contingency tables are defined on a grid point basis.

P SPATIAL-6: Interpreting the performance of higher resolution precipitation forecasts with spatial verification approaches

AUTHORS: Gofa, Flora; Boucouvala, Dimitra; Louka, Petroula; Flokas, Helena.

Presenting author e-mail: fgofa@hnms.gr; Affiliation (all authors): Hellenic National Meteorological Service, Athens, Greece

ABSTRACT: Although the resolution of numerical weather prediction models has increased remarkably, traditional grid-point-by-grid-point verification has not improved the quality of information for forecast performance or the relative characteristics of the derived fields. High-resolution numerical weather prediction (NWP) models produce more detailed precipitation structures but the real benefit is the more realistic statistics obtained from the higher resolution rather than the information for the specific grid point. Spatial verification methods in general reward closeness or resemblance by relaxing the requirement for exact matches between forecasts and observations. For instance, the neighborhood approach has the advantage of using a spatial window surrounding the forecast and/or observed points. The size of the neighborhood can be varied to provide verification results at multiple scales, enabling the determination of which scales the forecast has the most useful skill. On the other hand, feature-based spatial methods such as SAL enable insight into the attributes of forecast error in terms of the structure, amplitude or location characteristics of the precipitation fields. Apart from the different spatial methods that can be applied, it can be particularly interesting to compare traditional verification metrics with those produced from spatial approaches, and the assessment of the precipitation patterns that can be suitable for each application. A strong convective event is used as a test case for forecasting precipitation over the Mediterranean region. The available precipitation data are treated within both spatial and time windows using a variety of methods for averaging (upscaling), thresholding, and PDF generation as well as binary thresholded structures, each of which provid es distinctly useful information on model performance.

P SPATIAL-7: A Spatio-Temporal User-Centric Distance

AUTHORS: Brunet, Dominique; Sills, David; Casati, Barbara.

Presenting author e-mail: Dominique.Brunet@canada.ca; Affiliation (all authors): Environment and Climate Change Canada, Toronto, Canada

ABSTRACT: When predicting thunderstorms and localized severe weather events, "near misses" occur frequently and make obtaining representative verification results challenging. Indeed, a forecast close to a severe weather event would be counted as a false alarm, whereas not forecasting any severe weather would not properly represent the actual danger close-by. This is why the timing and location of the forecast relative to the impacted persons is one of the most important aspects that needs to be predicted. It allows the determination of a safe distance from the user for a given weather event forecast. Otherwise, action might need to be taken to minimize the possible impact of the weather event. In this study, we quantify the quality of a forecast in terms of spatial and temporal error relative to the location of the user. We propose a new forecast verification metric which is rigorous from a mathematical point of view, yet intuitive and user friendly in its interpretation, so that it could be used by outdoor event planners (concerts, sports, festivals), outdoor facility managers (campground, park, golf clubs), air traffic controllers, civil authorities as well as the general public. The proposed metric compares the distance from a location of interest to a forecast area against the distance from the same location to a set of observations via the Generalized Distance Transform. This is a generalization of the Hausdorff and Baddeley distances preserving metric properties while being robust and resilient to outliers. Time series for distances of forecasted and observed weather event relative to the user can then be constructed. This leads to the computation of the optimal time lag as a measure of timing error. Finally, the maximum difference between the time series can give a measure of spatial distance error for both the worst under-forecast and the worst over-forecast. An example application will be presented that compares lightning observations from the Southern Ontario Lightning Mapping Array (SOLMA) against the Canadian High-resolution Regional Deterministic Prediction System (HRDPS) at nominal resolutions of 2.5 km, 1 km and 250m.

P SPATIAL-8: A wavelet-based scale-separation verification approach to assess the added value of enhanced resolution models

AUTHORS: Casati, Barbara; Lussana, Cristian; Glazer, Anna; Milbrandt, Jason; Fortin, Vincent.

Presenting author e-mail: barbara.casati@canada.ca; Affiliation (all authors): Meteorological Research Division, Environment and Climate Change Canada, Dorval, Canada

ABSTRACT: High resolution Numerical Weather Prediction (NWP) systems produce precipitation forecast fields enriched of realistic small-scale details, as compared to their coarser-resolution driving models. Traditional point-by-point verification approaches, however, fails to reward the enhanced resolution, possibly due to the increased variability and small timing and location displacements. In this work we introduce a scale-separation verification skill score specifically defined to compare the performance of high resolution versus coarser resolution precipitation forecasts. Forecast and reference observation fields are decomposed into the sum of orthogonal wavelet components each characterized by a different spatial scale. The scale-dependence of the bias and the capability of the forecast to reproduce the observed scale structure are then assessed by comparing the wavelet component power spectrum. The scale dependence of the forecast accuracy and skill are assessed with the MSE and a MSE skill score (with reference = random), evaluated on the wavelet components for each spatial scale, separately. Note that the skill score so defined gives equal chance to high and lower resolution models, because the reference MSErandom (against which the model performance competes) is proportional to both observed and forecast variability (as opposed to the traditional MSE skill score, which uses as reference the climatology, which is proportional solely to the observed variability). The skill score is illustrated on the MesoVICT case studies, by comparing precipitation forecasts produced by the Canadian Global Environmental Multiscale (GEM) model, cascaded from the global (with 33km grid-spacing) to limited-area model domains with grid-spacing of 15 km and 2.5 km. The VERA analysis is used as the reference gridded-observation verification dataset. Timeseries of the scale-dependent scores provide useful insights in the evolution of the forecast performance, and the MesoVICT case studies are ranked according to the perfomance of the most prominent precipitation event.

P SPATIAL-9: A new spatial verification method of the gridded interpretations of T639 model products

AUTHORS: Wei, Qing.

Presenting author e-mail: weiqing@cma.gov.cn; Affiliation (all authors): National Meteorology Center of China, Beijing, China

ABSTRACT: From August 2015, the gridded products of interpretations of T639 model at 0.050 resolutions have been put into operational use at the National Meteorology Center of China. When verified using “point-to-point“ traditional method, they often score quite poorly because of the difficulty of predicting an exact match to the observations. To acquire a comprehensive assessment of the products, a new spatial verification method has been implemented. In this method, the forecast doesn‘t exactly match to the observation. More than 40000 automatic meteorological observational stations data in China are used as observation. Every forecast grid matches a range of observation stations within a radius of 10/20/30km, whichever station is at the same level with the grid, the verification result is correct. Also, the results are compared to that of “point-to-point“ traditional method.

P SPATIAL-10: Fuzzy Verification Test and Comparistion of Three Types of Severe Convective Weather Nowcasting

AUTHORS: Zhang, Xin; Li, Baiping; Dai, Jianhua; Mao, Mao.

Presenting author e-mail: jingshuiliushen784@126.com; Affiliation (all authors): Shanghai Central Meteorological Observatory, Shanghai, China

ABSTRACT: Severe convective weather is hard to forecast because of the characters of small scale and rapid development. Fuzzy verification methods developed these years can get evaluation information at different spatial scales by using a spatial window or neighborhood surrounding the forecast and/or observed points. We introduced five kinds of fuzzy verification methods (Upscaling, Minimum coverage, Fuzzy logic, Multi-event contingency table and Fraction skill score) into the verification work of severe convective weather of several Chinese Meteorological Observatory. Based on the one-hour reflectivity extrapolation products of Severe Weather Analysis and Nowcasting(SWAN) system of Chinese Meteorological Administration (CMA) and NowCAsting and Warning System(NoCAWS) of Shanghai Meteorological Bureau, we verified three kinds of severe convective weather (thunderstorm cell, squall-line and systemic heavy rainfall) with fuzzy verification methods. Furthermore, three ideal severe convective weather models are also built and verified to give a further study on the above-mentioned methods. The results show that: 1) Compared to traditional metrics with the strategy of "point to point", fuzzy verifications could give additional information in different scales and evaluation strategies, and evaluate forecasts more comprehensively and objectively. 2) Based on different evaluation strategies, one forecast has different optimal scales and each fuzzy verification method has its own feature and application. When forecast has large bias, fuzzy verification methods can still give effective or "useful" scores while traditional metric can only give poor scores. 3) The upscaling is a method with observations and forecasts both averaged to lager scales. The verification results tend to be better with a higher threshold and a larger neighborhood due to the special smoothing. However, isolated or severe convections could be decayed or missed with larger neighborhoods. Therefore, the upscaling method is applicable to large scale light precipitations. 4) The minimum coverage method asserts that A forecast is useful if the event is predicted over a minimum fraction of the region of interest?. The verification results tend to be better with a smaller minimum fraction. The "anywhere" standard is the most relaxed requirement It is easier to get high verification scores with lager neighborhood and we could get a perfect forecast(TS=1) if we allow a large enough neighborhood theoretically. The "anywhere" standard is more applicable to isolated or severe convections. For large scale precipitations the user need to choose a larger fraction standard and proper neighborhoods. However, we got better scores for squall-line and systemic heavy rainfall (typhoon Fitow) in our test while a low score for the verification for thunderstorm cell with the minimum coverage method. The poor performance of the one-hour reflectivity extrapolation products showed the great difficulty at predicting isolated heavy precipitations even with a relaxed requirem ent. 5) The fuzzy logic method has a wide application range from small scale to large scale precipitation. It is quite sensitive to the spatial distribution of rainfall and will get high scores at some scale neighborhoods for certain type of precipitations. Basically, traditional "good" forecasts could get high scores in small spatial-scale verification while "poor" forecasts could only get high scores in large spatial-scale verification. Therefore if a forecast gets a high score in small spatial-scale verification, the user should notice that it may be a good forecast. In the verification of three kinds of severe convective weather, we get some information with fuzzy logic method. The TS score decreased with larger threshold and neighborhood for the forecast of thunderstorm cell case; For the forecast of squall-line case the TS scores didn't show obvious differences with small neighborhood sizes, but the TS scores began to decreas when the neighborhood size was bigger than 33km; The stud y of typhoon Fitow as a respective of systemic heavy rainfall showed a similar result as that of the squall-line. 6) The Multi-event contingency table method is meaningful for the verification of isolated or severe convections. The results showed that the HK score decreased when neighborhood scale increased for low rainfall thresholds. However, the HK score had a positive correlation with neighborhood scale for high rainfall thresholds. 7) The FSS scores had a positive correlation with neighborhood scales and a negative correlation with rainfall threshold. The scores didn't obvious different rules in the verifications of three kinds of severe convective weather. 8) For the severe convective events with characteristics of high thresholds and small scales, the fuzzy verification methods including minimum coverage with low fraction, fuzzy logic and multi-event contingency table show more potential value than the traditional ones.

P SPATIAL-11: Scale-dependent evaluation of precipitation in UERRA regional re-analysis datasets for complex terrain regions

AUTHORS: Lussana, Cristian; Isotta, Francesco; Casati, Barbara; Frei, Christoph; Tveito, Ole Einar.

Presenting author e-mail: cristian.lussana@met.no; Affiliation (all authors): Norwegian Meteorological Institute, Oslo, Norway

ABSTRACT: Re-analysis aims at providing long-term, high-resolution, dynamically consistent climate datasets useful for environmental applications. Usually, these datasets covers a time period of several decades and they assimilate high-quality and detailed observations into the atmospheric analysis. In UERRA (Uncertainties in Ensembles of Regional ReAnalysis, FP7 research project of the EU), several regional deterministic and ensemble re-analyses, as well as downscaling procedures based on these re-analyses, are calculated for the past 30-50 years. With respect to precipitation, regional re-analyses are expected to better represents high threshold events than global re-analysis. The aim of the present study is to evaluate daily precipitation data from the new regional re-analyses of UERRA in two topographically complex sub-regions of Europe, namely the European Alps and Fennoscandia. Our focus is on aspects of re-analysis uncertainty that may be relevant for hydrological applications, notably the dependence on spatial scale. In comparison with verification of numerical weather predictions, the evaluation of reanalysis data must focus more on assessing the spatial smoothing and filtering properties of the atmospheric model considered, while the forecast error is less relevant because it is expected to be rather small due to the frequent data assimilation cycles. The evaluation is based on a comparison against gridded conventional climatological datasets derived from spatial analyses of high-resolution rain-gauge networks. The Alpine rain-gauge dataset, covers territories of seven countries and encompasses more than 5300 daily rain-gauge observations on average. In Fennoscandia, we focus on Norway, Sweden and Finland where a dataset of approximately 2000 daily rain-gauge observations is available. Our evaluation strategy combines "standard" verification measures of systematic and root mean squared deviations with spatial verification methods, which accounts for the weather field coherent spatial structure and presence of precipitation features. Scale dependence of the uncertainty/accuracy is examined in our analyses by considering (nested) hydrological catchments of variable size and by decomposing precipitation fields into (orthogonal) wavelets of variable scale. A difficulty for an unbiased evaluation of re-analyses is that reference datasets themselves are subject to uncertainties, the magnitude of which may be significant at the resolution of modern re-analyses and, hence, could affect a scale-dependent evaluation. This difficulty is addressed in the Alpine section of our analysis by introducing a new probabilistic rain-gauge dataset which explicitly quantifies uncertainties by ensembles.

P SPATIAL-12: Spatial verification of precipitation forecasts over Korea by the KIAPS model

AUTHORS: Lee, Juwon; Lee, Eun-Hee.

Presenting author e-mail: jw.lee@kiaps.org; Affiliation (all authors): Korea Institute of Atmospheric Prediction Systems (KIAPS), Seoul, Korea

ABSTRACT: Rainfall forecasts can be verified a number of ways. Traditionally, several statistical scores can be used diagnosis of quantitative precipitation forecasts (QPFs) errors, but these methods exhibit some problems, including double-penalty and domination of score by small-scale features. So, we need new spatial verification techniques which can account for field spatial structure and uncertainties in location and provide information on error in physical terms. Among them, Method for Object-based Diagnostic Evaluation (MODE) represents a class of a spatial varication methods is to identify localized features of interest in scalar fields and compare features in two fields to identify which features best correspond to each other. Thus, in this study, the systematic errors of the next generation global model under developed by Korea Inst itute of Atmospheric Prediction Systems (KIAPS) since 2011 (KIM, KIAPS Integrated Model) were estimated using MODE in which the forecast rain pattern correlated well with the observed rain pattern. This result can be used for improvement of KIM.

P SPATIAL-13: Calculus-A Novel Quality Measure for the Verification of High Resolution Quantitative Precipitation Forecasts

AUTHORS: Huang, Juan; Zhao, Fang.

Presenting author e-mail: huangmen_1984@sohu.com; Affiliation (all authors): Zhejiang Meteorological Observatory, hangzhou, china

ABSTRACT: A novel Fuzzy verification measures,which contains three aspects of the structure(S),amplitude(A) and location(L) of the precipitation grids that combined with a variable precipitation window is introduced for the verification of the high resolution forecasts based on grids. Compared to traditional metrics with the stagey of “point to point“,calculus methods can give the form error of the precipitation field.Compared to object-based quality measure, it is more visualer and easier.Focusing on three types of severe convection weather(e.g.,an upper trough type ,subtropical high edge type and east-wave type),calculus methods is applied to the verification of one-hour radar extrapolation rain product which space resolution is 0.01°×0.01°.The result showed that the S,A, and L values of Sudden enhancement e cho was much larger than stability echo in upper trough type, larger values of the massive echo new bored or into the sea from the land in subtropical high edge type and larger values of typhoon peripheral spiral cloud rainfall in east-wave type contrast to typhoon rainfall.The speed of echo extrapolation can be adjusted to reduce the error when L is fairly small.Calculus methods is applied to one-hours short time QPF verification which contained radar and other fusion data from March to October, 2016 . In the first hour has the Minimum values of the S,A, and L, meantime in the second hour, the value is increasing with time, but not increase between the third hour and sixth hour.

P SPATIAL-14: Verification of Rainfall Spatial Forecast over India: A method to account for Location Errors

AUTHORS: Sharma, Kuldeep; Ashrit, Raghavendra; Bhatla, Rajeev; Mitra, A.K.; Iyengar, Gopal; Rajagopal, E.N..

Presenting author e-mail: kuldeep@ncmrwf.gov.in; Affiliation (all authors): National Centre for Medium Range Weather Forecasting, Noida, India

ABSTRACT: In recent years, forecasting skill of numerical weather prediction (NWP) models has been improved tremendously. This improvement is attributed to increased resolution, global advanced data assimilation and satellite observations. NCMRWF also updates its numerical weather prediction modeling system with UK Met office. Two state-of-art NWP models have been investigated over the India to assess their ability during the monsoon season 2016. The unified model operational at National Center for Medium Range Weather Forecasting (NCUM) and the unified model operational at the Australian Bureau of Meteorology (Australian Community Climate and Earth-System Simulator Global (ACCESS-G)) are used in this study. Traditional verification at each grid point may often indicate poor performance because of the increased small-scale variability so that the true quality of the forecasts is not always characterized well. Very often, the models correctly forecast the high rainfall amounts associated with severe weather events. At times, due to slight miss match in the location or/and pattern of the actual event, the forecast verifications yield poor scores. In some of the applications, the model forecasts of heavy rains are of great value even with errors in location and patter. For example, the rainfall forecast over a catchment or a river basin which could be used to estimate the river runoff and discharge downstream. It is important to evaluate the rainfall forecasts keeping in mind the applications, like Hydrometeorology and flood forecasting. This study adopts a novel strategy of verification to accommodate the predicted rains in the neighborhood of actual events. Using this approach, contingency analysis is performed for categorical verification. Standard verification metrics like Bias, Equitable Threat Score (ETS), Probability of Detection (POD) and HK-scores etc have been examined during the monsoon season 2016 from Day-1 to Day-5 forecasts. Along with the traditional verification scores, a new family of scores like EDS (Extreme Dependency Score), EDI (Extremal Dependence Index) and Symmetric EDI (SEDI) are also computed to examine the skill of forecast of NWP models in case of extreme rainfall during the season.

P SPATIAL-15: Spatial verification of high-resolution precipitation forecasts through wavelets

AUTHORS: Kapp, Florian; Friederichs, Petra; Weniger, Michael.

Presenting author e-mail: fkapp@uni-bonn.de; Affiliation (all authors): Meteorological Institute, University of Bonn, Bonn, Germany

ABSTRACT: Wavelet transforms provide a framework to analyze spatial data on separate scales and in different directions. Several wavelet based spatial verification techniques have been developed over the past decade, which focus on the assessment of forecast performance on distinct physical scales (e.g. spatial or small-scale precipitation). However, wavelet transforms as advanced in scientific fields such as texture analysis, face or fingerprint recognition or feature detection still reveal unused potential in meteorological applications. We present a technique developed for image texture analysis to evaluate spatial forecasts of precipitation. The forecasts are provided by the the high-resolution COSMO-DE ensemble prediction system of the German meteorological service (Deutscher Wetterdienst). A discrete wavelet transform is used to estimate the forecasts power-spectra on distinct scales. We thus concentrate on the verification of spatial structures of precipitation instead of a gridpoint-by-gridpoint comparison. A linear discriminant analysis is applied to analyze ensemble member clustering. Image texture analysis shows promising results in discriminating ensemble members by their distinct driving boundary conditions, as well as different forecast dates.

P OBS UNC-1: Quantifying uncertainty in rainfall datasets over Africa

AUTHORS: Lennard, Christopher.

Presenting author e-mail: lennard@csag.uct.ac.za; Affiliation (all authors): University of Cape Town, Cape , South Africa

ABSTRACT: Rainfall-based studies use some form of observed rainfall data to e.g. document historical rainfall characteristics, develop a baseline rainfall climatology against which future change can be compared, quantify current or projected water availability to assess rainfall driven activities, evaluate rainfall driven impacts and potential future changes in rainfall may have on these. Observations of rainfall (data) can be obtained from a number of different sources: rain gauges at measuring stations report locale-specific rainfall, remotely sensed rainfall data from terrestrial and satellite-based observation stations deliver rainfall data at a wider spatial scale; reanalysis rainfall data that is a product of a climate model and finally some rainfall data sets are a combination of one or more of the above rainfall products to produce a final multi-sourced rainfall product. The most popular reference rainfall data in used climate literature are the CRU, GPCP, GPCC, TRMM and UDEL datasets. However, there are many other rainfall datasets available for consideration, for example, CMORPH, FEWS, TAMSAT & RIANNAA, TAMORA and the WATCH & WATCH-DEI data. This paper presents a number of rainfall datasets available over Africa and assesses the spatial and temporal differences that exist between these. Taylor diagrams are used to quantify these differences based on homogenous rainfall regions and spatial maps to highlight hotspots of regional uncertainty. Reasons for the differences are presented. A final recommendation is made that in the context of the quantified observational uncertainty presented that a probabilistic view of historical rainfall should be adopted in rainfall studies over Africa.

P OBS UNC-2: Effects of observation uncertainties in verification practices

AUTHORS: Casati, Barbara; Lemay, Francois; Husson, Thierry; Fortin, Vincent.

Presenting author e-mail: barbara.casati@canada.ca; Affiliation (all authors): Meteorological Research Division, Environment and Climate Change Canada, Dorval, Canada

ABSTRACT: Observations can be affected by different types of uncertainties, such as measurement errors (e.g. due to instrument failure), representativeness and/or sampling errors, uncertainties inherent in the assumptions of remote-sensing retrieval algorithms (e.g. for satellite or radar-borne observations) and uncertainties introduced by gridding procedures. The effects of these observation uncertainties on verification results are yet to be fully quantified, and how to account for them in verification practices is still an open research topic. This presentation illustrates some of the issues encountered in the Canadian operational verification environment due to the presence of observation uncertainty, with the aim of identifying the observation uncertainties which have largest impacts on verification results.

P CLIM-1: Contribution of Black Carbon induced Climate Change on mountain snow-melt over Eastern Africa region

AUTHORS: Mbithi, Daniel; Muthama, Prof. Nzioka; Ininda, Prof. Joseph; Onwnga, Prof. Richard.

Presenting author e-mail: mbithi_d@yahoo.com; Affiliation (all authors): University of Nairobi, Nairobi, Kenya

ABSTRACT: This study aims to assess effect of land-based black carbon (BC) aerosols on Climate Change using radiative forcing (RF). Also, modeling of future rainfall and temperature change over Eastern Africa region will be performed. Data to be used include Coordinated Regional climate Downscaling Experiment (CORDEX) data spanning 1950-2016; Clouds and Earth‘s Radiant Energy System (CERES) satellite data on radiative forcing at top of atmosphere (TOA) spanning 2000-2016. Further, the Moderate Resolution Imaging Spectroradiometer (MODIS) data spanning 2000-Present on land surface reflectance, emissivity, albedo, land cover characteristics including thermal anomalies over Eastern Africa region will be used. Spatial-Temporal Characteristics of the direct radiative forcing of BC will be determined. The former will be carried out using surfer software while the later will be determined statistically and output determined by time series analysis. Principal Component Analysis (PCA) will be done between the main Climatic parameters (Rainfall and Temperature) and RF (ATM) estimates. Contribution of (RF) ATM to global warming will be estimated using the IPCC net RF values as reference. The estimates will be mapped over the area of study. Future global warming will be estimated over the area of study using Model for Assessment of Green House gas Induced Climate Change: A Regional SCENario GENerator (MAGICC SCENGEN). It is expected that radiative forcing due to black carbon over Eastern Africa region will be determined and past and future precipitation and temperature scenarios generated. Results of this study will provide information on black carbon induced radiative forcing estimates the region of study, radiative interactions and future climate scenarios.

P CLIM-2: Calibration of decadal ensemble predictions

AUTHORS: Pasternack, Alexander; Rust, Henning W.; Grieger, Jens; Bhend, Jonas; Liniger, Mark A.; Ulbrich, Uwe.

Presenting author e-mail: alexander.pasternack@met.fu-berlin.de; Affiliation (all authors): Freie Universität Berlin, Berlin, Germany

ABSTRACT: Decadal climate predictions are of great socio-economic interest due to the corresponding planning horizons of several political and economic decisions.To capture uncertainties (e.g. due to initial condition uncertainty) weather and climate predictions are issued in a probabilistic way. One issue frequently observed for probabilistic forecasts is that they tend to be not reliable, i.e. the forecasted probabilities are not consistent with the relative frequency of the associated observed events. Thus, these kind of forecasts need to be re-calibrated. While re-calibration methods are frequently applied for seasonal time scales, these methods still have to be adapted for decadal time scales and its characteristic problems like climate trend and lead time dependent bias. Here, we propose a method to re-calibrate decadal ensemble predictions that takes the above mentioned characteristics into account. The required parameters of this method are estimated by numerical optimization of a verification score. In this regard, this method will be applied to decadal forecasts from the MiKlip system (Germany‘s initiative for decadal prediction) and validated with respect to different optimization scores.

P CLIM-3: Challenges in the user-oriented verification of wind speed climate predictions

AUTHORS: Torralba, Verónica; Doblas-Reyes, Francisco J.; Cortesi, Nicola; Marcos, Raül; Lledó, Llorenç; Gonzalez-Reviriego, Nube.

Presenting author e-mail: veronica.torralba@bsc.es; Affiliation (all authors): Barcelona Supercomputing Center (BSC), Barcelona, Spain

ABSTRACT: Climate forecast verification tailored to the wind energy sector aims to quantify the adequacy of the forecast uncertainty estimates, but also to increase the trustworthiness of these users in the potential added value of these predictions for some particular applications such as the cash flow anticipation or the maintenance planning. The assessment of the probabilistic wind speed climate predictions for wind energy applications can be affected by deficiencies in the forecasts and in the reference datasets. In this assessment, one of the challenges is the unavailability of wind speed observational measurements with time series long enough to produce a robust evaluation of the climate predictions at global scale. This limitation makes the forecast verification based on real observations difficult and suggests the use of reanalysis pr oducts as a reference for the forecast verification process. However, the quality of wind speed from global reanalyses is linked to the different methodologies the reanalyses use to infer 10-metre wind speed from the lowest model level and variations in the observational sources used in the assimilation process. A comprehensive comparison of the seasonal climatological wind speed probability distribution moments in three state-of-the-art atmospheric reanalyses (ERA-Interim, JRA-55, MERRA-2), and the forecasts produced by the ECMWF System 4 seasonal climate prediction system has been performed. Results show important disagreements between the reanalyses, particularly for JRA-55 that overestimates long-term trends over land. This disagreement should be considered by users who employ reanalysis data for the evaluation of the long-term wind speed variability. Beyond the uncertainty in the references, climate forecast systems show also a range of challenges for the efficient use of clima te predictions such as their low level of predictability in some regins, important systematic errors, and the lack of reliability. All of them need to be considered for their application by the industry. As an example, extratropical weather regimes produced by the prediction systems and the reanalyses have been compared to explore the possible causes of the low predictability, which could be attributed to the misrepresentation of some types of atmospheric flows by the climate forecast systems. In addition, we have proposed three bias adjustment methods to ensure the usability of the forecasts: simple bias correction, calibration and quantile mapping. The predictions resulting from this post-processing show improved forecast quality and address the requirements of the users.

P S2S-1: Evaluation of Heat Wave forecasting using large scale circulation pattern based on TIGGE data

AUTHORS: Lee, Hyun-Ju; Lee, Woo-Seop; Yoo, Jin-Ho.

Presenting author e-mail: asteria1104@apcc21.org; Affiliation (all authors): Climate Prediction Team, Climate Research Department, APEC Climate Center, Busan, Republic of Korea

ABSTRACT: This study assesses the predictability of heat wave occurrence over South Korea using heat waves index associated with large-scale circulation by applying operational medium-range ensemble forecast datasets available from The Observing system Research and Predictability Experiment Interactive Grand Global Ensemble (TIGGE) portal. The occurrence of heat waves over Korea was forecasted and verified for the period between July and August of 2013 using the TIGGE models under the three methods that are listed as follows: (a) TMAX shows forecast of heat waves when the daily maximum temperature exceeds the upper 95th percentile value of the observed climatological PDF, (b) BCT indicates forecast of heat waves when the daily maximum temperature exceeds the value of the upper 95th percentile of each model‘s climatological PDF and (c) HWI shows forecast of heat waves when the Heat Wave Index, which is based on large-scale circulation, is positive. Here, HWI is defined as difference in the 200 hPa vorticity between the average over 25~30N, 110~130E and the average over 35~45N, 120~140E. To evaluate the performance of the heat wave forecast, the percent correct (PC), threat score (TS), and equitable threat score (ETS) were employed from the two-by-two contingency table. When using HWI, showed higher predictability of heat waves than those of maximum temperature (TMAX) and Bias corrected TMAX. The verification scores using HWI were highest compared to other cases for day 5 through day 9 forecasts. It is concluded that the proposed HWI can provide reliable information on heat waves in advance and provide more prevention time for the public in South Korea. Also it can be useful for heat wave forecasting which could reduce the health impacts of heat waves through appropriate and timely mitigation efforts.

P S2S-2: Verification for long-lead station-scale prediction of hydrological droughts in South Korea based on bivariate pattern-based downscaling

AUTHORS: Sohn, Soo-Jin; Tam, Chi-Yung.

Presenting author e-mail: jeenie7@apcc21.org; Affiliation (all authors): APEC Climate Center (APCC), Busan, Republic of Korea

ABSTRACT: Capturing climatic variations in boreal winter to spring (December to May) is essential for properly predicting droughts in South Korea. The variability and predictability of the South Korean climate during this extended season, based on observations from 60 station locations and multi-model ensemble (MME) hindcast experiments (1983/84-2005/26) archived at the APEC Climate Center (APCC), is investigated. This study verifies the performance of precipitation and temperature predictions based on a perfect empirical model and a hybrid dynamical-statistical model, and that for extreme drought predictions. Multivariate empirical orthogonal function (EOF) analysis results based on observations show that the first two leading modes of winter-to-spring variability, which together account for ~80% of the total variance, are characterized by regional-scale precipitation and temperature anomalies covering the whole South Korea territory. These modes were also closely related to some of the recurrent large-scale circulation changes in the northern hemisphere during the same season. Consistent with the above, examination of the standardized precipitation evapotranspiration index (SPEI) indicates that drought conditions in South Korea tend to be accompanied by regional-to-continental-scale circulation anomalies over East Asia to the western north Pacific. Motivated by the aforementioned findings on the spatial-temporal coherence among station-scale precipitation and temperature anomalies, a new bivariate and pattern-based downscaling method was developed. The novelty of this method is that precipitation and temperature data were first filtered using multivariate EOFs to enhance their spatial-temporal coherence, before being linked to large-scale circulation variables using canonical correlation analysis (CCA). To test its applicability and to investigate its related potential predictability, a perfect empirical model was first constructed with observed datasets as predictors. Next, a model output statistics (MOS)-type hybrid dynamical-statistical model was developed, using products from nine one-tier climate models as inputs. The final station-scale temperature and precipitation were then used to produce SPEI or drought forecasts. Temporal correlation as well as the linear error in probability score (LEPS), which measures the error in the probability space rather than in the real measurement space, were used to assess the skill of extreme predictions. It was found that, with model sea-level pressure (SLP) and 500hPa geopotential height (Z500) as predictors, statistically downscaled MME (DMME) precipitation and temperature predictions were substantially improved compared to those based on raw MME outputs. Limitations and possible causes of error of such a dynamical-statistical model, in the current framework of dynamical seasonal climate predictions, were also discussed. Finally, the downscaling method was used to construct a dynamical-statistical system for 6 month-lead drought predictions for 60 stations in South Korea. DMME was found to give reasonably skillful long-lead forecasts of SPEI for winter to spring. Moreover, DMME-based products clearly outperform the raw MME predictions, especially during extreme wet years. Our results could lead to more reliable climatic extreme predictions for policymakers and stakeholders in the water management sector, and for better mitigation and climate adaptations.

P S2S-3: MJO verification by JMA Global Ensemble Prediction System

AUTHORS: Kubo, Yutaro.

Presenting author e-mail: yutaro.kubo@met.kishou.go.jp; Affiliation (all authors): Japan Meteorological Agency, Tokyo, Japan

ABSTRACT: Madden-Julian Oscillation (MJO) is the dominant mode of intra-seasonal variability in the tropics. MJO influences not only the tropics but also the extratropical circulations. Thus, it is important to predict MJO and its response. This study investigates MJO forecast skill and that performance with the JMA Global Ensemble Prediction System (GEPS), using the GEPS re-forecast dataset by the algorithm of Wheeler and Hendon (2004). MJO of GEPS is verified against JRA-55 reanalysis data, with using the diagnostic package developed by the U.S. Climate Variability and Predictability (CLIVAR) MJO Working Group. The correlation coefficient of MJO is above 0.6 until a lead time of about 12days. The tropical circulation in response to MJO and the extratropical circulation by that teleconnection mechanism are also reproduced until a lead time of about 2 weeks in terms of composite analysis. In phase 2-3 (the active convection by MJO is located near the Indian Ocean), MJO phase speed is a little faster than analyzed phase speed. MJO amplitude is also a little smaller than analyzed amplitude. Seeing phase 2-3 composite map, the positive precipitation anomaly by MJO is week, but upper divergence anomaly is reproduced.

P S2S-4: Verification and predictability of the surface heat and freshwater fluxes on seasonal and decadal timescales using multiple reanalysis data sets

AUTHORS: Romanova, Vanya; Wahl, Sabrina; Hense, Andreas.

Presenting author e-mail: romanova@uni-bonn.de; Affiliation (all authors): University of Bonn, Bonn, Germany

ABSTRACT: Standard verification in meteorological forecasts usually considers temperature, geopotential heights, pressure or precipitation, which are directly measured variables. However, of strong interest are the large scale air-sea energy fluxes since they play important role in climate variability. These variables are composite variables and their uncertainties are defined as a sum of the uncertainties of all consisting terms. Their forecast seems a challenging task especially when considering longterm timescales. An attempt is made to predict the surface fluxes retrospectively through ensemble hindcasts using different methods for ensemble generation. Evaluation of the hindcasts on independent observational data sets and different reanalysis products is conducted on the following evaluation metrics: 1) box plots, lower and upper whiskers and the outliers; 2) RMS of the perturbation patterns; 3) anomaly correlation coefficients (ACC); 4) Ensemble Spread Skill Score (ESS); 5) cumulative frequency distribution (probability estimates), across all grid points for single hindcasts and single observation data set; and 6) reliability diagrams. The evaluation of the hindcasts showed reasonable spreads during the whole integration period, but still too low compared to the spread of the ensemble of reanalysis. To calibrate the outcomes from the model predictions and the observations, the difference in the description of the uncertainties should be considered.

P S2S-5: The TIGGE and S2S Museums - websites of ensemble forecast products

AUTHORS: Matsueda, Mio; Nakazawa, Tetsuo; Yamagami, Akio.

Presenting author e-mail: mio@ccs.tsukuba.ac.jp; Affiliation (all authors): University of Tsukuba/University of Oxford, Tsukuba/Oxford, Japan/UK

ABSTRACT: We introduce two websites displaying a variety of ensemble forecast products: the TIGGE (The International Grand Global Ensemble) and S2S (Subseasonal to Seasonal) Museums (http://gpvjma.ccs.hpcc.jp/TIGGE/ and http:// gpvjma.ccs.hpcc.jp/S2S/). The TIGGE Museum displays various products of 10 global medium-range ensemble (TIGGE) forecasts, originally provided by the WWRP THORPEX project to accelerate improvements in the accuracy of one-day to two-week high impact weather forecasts for the benefit of society, the economy and the environment. The TIGGE Museum includes statistical verifications of TIGGE forecasts, ensemble-based probabilistic forecasts of severe weather, and forecasts of the Madden-Julian Oscillation (MJO), atmospheric blocking, and teleconnection pattern indices. The S2S Museum displays various products of 11 global S2S ensemble forecasts and reforecasts, provided by the WWRP/THORPEX/WCRP joint S2S project to improve forecast skill and understanding on the S2S timescale, and promote its uptake by operational centres and exploitation by the applications community. The S2S Museum contains forecasts of the Arctic/Antarctic Oscillation (AO/AAO) indices, the North Atlantic Oscillation (NAO) index, teleconnection pattern indices (the Pacific/North American (PNA), the Western Pacific (WP), and the Eurasian (EU) pattern indices), wave activity flux at 200 hPa, the Sudden Stratospheric Warming (SSW), MJO, sea surface temperature, and sea-ice cover. The forecast products available at the TIGGE and S2S Museums are regularly updated every day, with a 2- and 21-day delay, respectively, and are available only for research and education purposes.

P S2S-6: Assessment of WMO LC-LRFMME Models in Seasonal Climate Prediction

AUTHORS: Kim, Gayoung; Ahn, Jong-Bae; Kryjov, Vladimir N.; Sohn, Soo-Jin; Graham, Richard; Kolli, Rupa Kumar.

Presenting author e-mail: chorong@apcc21.org; Affiliation (all authors): APEC Climate Center, Busan, Korea

ABSTRACT: The World Meteorological Organization (WMO) Lead Centre for Long-range Forecast Multi-model Ensemble (WMO LC-LRFMME) has been established to collect and share long-range forecasts from the WMO designated Global Producing Centers (GPC). In this study, the seasonal skill of the deterministic multi-model prediction of GPCs in WMO LC-LRFMME is investigated. The GPC models included in the analysis cover 30 years of common hindcast period from 1981 to 2010 and real-time forecast for the period from DJF2011/2012 to SON2014. The equal-weighted multi-model ensemble (MME) method is used to produce the MME forecast. We show that the GPC models generally capture the observed climatological patterns and seasonal variations in temperature and precipitation. However, some systematic biases/errors in simulation of the climatological mean patterns and zonal me an profiles are also found, most of which were located in mid- or high-latitudes. The temporal correlation coefficients both of 2m temperature and precipitation in the tropical region (especially over the ocean) exceed 95%, but drop gradually toward high-latitudes and are even negative in the polar region for precipitation. The prediction skills of individual models and the MME over 13 Regional Climate Outlook Forum (RCOF) regions for four calendar seasons are also assessed. The prediction skills vary with season and region, with the highest skill being demonstrated by the MME forecasts for the regions of the tropical RCOFs. These predictions are strongly affected by the ENSO over Pacific Islands, Southeast Asia and Central America. Additionally, Southeast of South America and North Eurasian regions show relatively low skills for all seasons when compared to other regions.

P S2S-7: A verification framework for South American sub-seasonal precipitation predictions

AUTHORS: Coelho, Caio; Firpo, Mári.

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

ABSTRACT: Sub-seasonal precipitation predictions - here referred to as precipitation predictions for weekly periods produced one to four weeks in advance - are relevant for strategic decisions and planning in various South American economic sectors (e.g. agriculture, water management and hydropower generation). The recent availability of these predictions produced within the context of the joint World Weather Research Program (WWRP)/World Climate Research Program (WCRP) Sub-seasonal to Seasonal (S2S) prediction project allows the investigation of the current skill level of the participating modeling centers. However, a verification strategy is required in order to document both the deterministic and probabilistic predictive skill in support of possible future routine sub-seasonal predictions as verification information detailing past mo del performance is a key prediction practice component to enhance forecasters´ confidence. This study proposes a verification framework for this purpose. Ensemble sub-seasonal precipitation retrospective predictions (hindcasts) for the period 1999-2010 and near real time ensemble predictions issued in 2015 and 2016 from the European Centre for Medium-Range Weather Forecasts (ECMWF) are used for illustrating the proposed verification framework, which includes the following procedures: a) comparison of the predicted (ensemble mean) and observed precipitation anomaly pattern for a selected case study, with the strength of correspondence measured with spatial pattern correlation; b) construction of maps showing the correlation between the predicted ensemble mean and observed precipitation anomalies at each grid point, with the aim of assessing the strength of the linear association between the predicted and observed anomalies; c) use of ensemble predictions for constructing maps showing the area under the Relative Operating Characteristic (ROC) curve for probabilistic predictions for the event positive precipitation anomaly issued at each grid point, with the aim of assessing discrimination ability (i.e. ability to successfully distinguish events from non-events); d) construction of reliability diagrams for ensemble derived probabilistic predictions issued for the event positive precipitation anomaly, with the aim of assessing reliability (i.e. how well calibrated the issued probabilities are) and resolution (i.e. how the frequency of occurrence of the event differs as the issued probability changes). Both correlation and ROC maps as well as the reliability diagrams are constructed using both hindcasts and the two years of near real time predictions to allow comparisons. Regions consistently presenting favorable skill through the computed verification scores, particularly in terms of association and discrimination attributes, are identified as where sub-seasonal p redictions are most likely to be successful. The proposed framework is useful not only in support of sub-seasonal prediction practice, but also to provide feedback to model developers in identifying strengths and weaknesses for future sub-seasonal predictions systems improvements.

P MISC-1: The Impact of Moisture Flux and Vertical Wind Shear on Mesoscale Convective Systems Over Nigeria

AUTHORS: Olaniyan, Olumide; Ajayi, Vincent.

Presenting author e-mail: olaniyan_olumide@yahoo.co.uk; Affiliation (all authors): Nigerian Meteorological Agency, Abuja, Nigeria

ABSTRACT: This study is an effort to understand heavy rainfall events over Nigerian during the summer months of June, July, August and September (JJAS) of year the 2010 to 2014. This is aimed to determine the Impact of moisture flux and vertical wind shear on mesoscale convective systems (MCSs) over Nigeria. Other derived meteorological parameters such as wind divergence, convective available potential energy (CAPE) and equivalent potential temperature are also assessed to ascertain the threshold values of these parameters responsible for isolated cases of extreme precipitation in order to enhance extreme rainfall predictability. The ECMWF reanalysis (0.125° x 0.125°) and TRRM of (0.25° x 0.25°) datasets were utilized for this purpose. The values of the above mentioned derived parameter during the storm are analysed using Ferret analytical package and superimposed on the precipitation at the surface obtained from Satellite dataset to determine areas of highs and lows and compared with the corresponding values of rainfall. The mean values of the derived parameters three days prior and after the storms are also obtained to assess their features before, during and after the storm. The result shows that horizontal Moisture flux divergence (HMFD) values range between 0.50 and $2.0x10^{-6}gKg^{-1}s^{-1}$ at the surface is capable of producing substantial amount of rainfall mostly over 50mm. The CAPE analysi indicated that CAPE above $1500Jkg^{-1}$ is favourable for convection over the northern parts, while less than 1500Jkg-1 can trigger convective activities over the southern part. The mean moisture flux, CAPE and saturated potential parameters start to accumulate few days before the storm and peak on the day of the convective storms, then starts declining after the storm. The result also shows that sufficient CAPE and wind shear is not enough for convection, it is necessary to have sufficient moisture to sustain the storm through its life.

P MISC-2: Empirical statistical modeling of March-May rainfall prediction over southern nations, nationalities and people‘s region of Ethiopia

AUTHORS: Hailesilassie, Wondimu; Tsidu, Gizaw.

Presenting author e-mail: wonde721@gmail.com; Affiliation (all authors): Rift Valley Lakes Basin Authority, Hawassa, Ethiopia

ABSTRACT: Statistical predictive models were developed to investigate how global rainfall predictors relate to the March-May (MAM) rainfall over Southern Nations, Nationalities and People‘s Region (SNNPR) of Ethiopia. Data utilized in this study include station rainfall data, oceanic and atmospheric indices. Because of the spatial variations in the interannual variability and the annual cycle of rainfall, an agglomerative hierarchical cluster analyses were used to delineate a network of 20 stations over study area into three homogeneous rainfall regions in order to derive rainfall indices. Time series generated from the delineated regions were later used in the rainfall/teleconnection indices analyses. The methods employed were correlation analysis and multiple linear regressions. The regression modes were based on the training pe riod from 1987-2007 and the models were validated against observation for the independent verification period of 2008-2012. Results obtained from the analysis revealed that sea surface temperature (SST) variations were the main drivers of seasonal rainfall variability. Although SSTs account for the majority of variance in seasonal rainfall, a moderate improvement of rainfall prediction was achieved with the inclusion of atmospheric indices in prediction models. The techniques clearly indicate that the models were reproducing and describing the pattern of the rainfall for the sites of interest. For the forecast to become useful at an operational level, further development of the model will be necessary to improve skill and to determine the error bounds of the forecast.

P MISC-3: Application of a Bias Correction Scheme for 2 Meter Temperature in Numerical Weather Forecast Model

AUTHORS: Zhao, Bin; Zhang, Bo.

Presenting author e-mail: zhaob@cma.gov.cn; Affiliation (all authors): National Meteorological Center, beijing, china

ABSTRACT: The inherent difference between the observation topography and model terrain has seriously affected the 2 m temperature verification accuracy. The traditional two-dimensional interpolation scheme can only ensure forecast element and observation consistency in latitude and longitude location of the two-dimensional space, while ignoring the vertical direction consistency, which makes verification result of forecast and observation is not from the same position, thereby causing an evaluation misleading. Diurnal cycle is an important feature of the 2 m temperature, Due to the limitation of physical process such as radiation, large bias always appears in the diurnal cycle forecast. In this paper, using three-dimensional forecast variables, combined with near-surface elements of forecast products, an advanced three-dimensional interpolation scheme is developed to ensure consistency with the observed three-dimensional space forecasting. Based on the topography correction methods, monthly forecast error is used as reference bias products to eliminate systematic error and obtain forecast products with characteristics of diurnal cycle. Based on typical observation gauges in Shanxi province, 48 hours forecast products on August 2016 are used to compare and it is found that the three-dimensional interpolation scheme effectively solve the evaluation misleading caused by the height bias between the model terrain and observation topography, however it can hardly effectively improve the diurnal cycle trend. After the systematic error correction is adopted, the diurnal cycle forecasting features are improved obviously. In particular, it shows better consistency with observation and the higher skill score, especially in the first 24 hours. By monthly statistical evaluation, it indicates that 2 m temperature after the bias correction can effectively improve the oscillation of periodic errors, the RMSE keeps at about 2k which shows the obvious advantage of improvement.

P MISC-4: A study on evaluation method for predictability of persistent heavy rainfall events over East Asia based on ensemble forecast

AUTHORS: Wu, Zhipeng; Chen, Jing; Zhang, Hanbin.

Presenting author e-mail: 361913145@qq.com; Affiliation (all authors): Chongqing Meteorological Observatory, Chongqing City, China

ABSTRACT: The Persistent Heavy Rainfall (PHR) is the most influential extreme weather event in Asian summer monsoon season, which has attracted intensive interests of many scientists. By use of operational global ensemble forecasts from China Meteorological Administration(CMA), a new verification method applied to evaluate the predictability of PHR is investigated. A metrics called Index of Composite predictability (ICP) based on Equitable Threat Score(ETS) of 24h accumulated precipitation and Root Mean Square Error(RMSE) of Height at 500hPa is established to identify “good“ and “poor“ forecast from ensemble members. Using the metrics of ICP, the predictability of two typical PHR events in June 2010 and June 2011 is estimated. The results show that the “good member“ and “poor member“ can be identified in the context of ICP and present an obvious discrepancy in predicting the key weather system impact on PHR. The different performance of “good member“ and “poor member“ reveals the higher predictability both in synoptic scale and mesoscale weather system in their location, duration and the movement by “good member“. The source of growth errors for “poor“ member is mainly from errors of initial conditions in northern polar region. Keywords: persistent heavy rainfall, predictability, ensemble prediction, index of composite predictability, source of growth error.

P MISC-5: The Analysis of the Predictability of an Extreme Rainfall Event in July 18-21,2016

AUTHORS: Ma, Jie.

Presenting author e-mail: maj@cma.gov.cn; Affiliation (all authors): National Meteorological Center, CMA, Beijing, china

ABSTRACT: Using the products of the European Centre for Medium Range Weather Forecasting (ECMWF) ensemble and deterministic model, this research discusses the predictability of an extreme rainfall event occurring in 18-21 July. The result shows the character of this rainfall event is the evident variability of the strength and the position of rainfall within a short time occurring around July 19th, especially the position of the rainbelt which moved northward almost 10 degree during the period from 18 July to 19 July. So the substantial challenge encountered when operational meteorologists faced the northern edge of heavy rainfall, and which caused the heavy precipitation of North China was poorly predicted. Further analyses reveal that the sensitivity of the Northern edge of heavy rainfall stems from the sensitivity of the str ength of Pohai blocking high system. While there appears a block high system above Pohai region, the vortex in Northern China will move more slowly northward and get the higher latitude. Then the interaction between the vortex and the block high will strengthen the south-west wind in low levels, the rainbelt intensifies and shifts farther northward. Furthermore, comparisons the performance between the ensemble and deterministic model in different leading time, It demonstrates that the maximum limitation of predictability for the location of rainbelt is below 96 hours.

P MISC-6: Verification of 10-metre wind-speed forecast and error analysis in China

AUTHORS: Li, JiaRui; Dong, Quan; Zhang, Kun.

Presenting author e-mail: lijr@cma.gov.cn; Affiliation (all authors): National Meteorological Center of China Meteorological Administration, Beijing/Haidian, People‘s Republic of China

ABSTRACT: In this paper, the accuracy of 10-metre wind-speed forecast is verified by using observed data after a statistical quality control procedure, then error characteristics are analyzed to realize the error‘s spatial distribution and its influence factors in operational forecast. Results show that when the actual wind speed is less than 3.4m/s, forecast derived from ECMWF numerical prediction model is higher than the observed data, but turn to be lower while the actual wind speed exceeds 5.5m/s. The deviation increases not only with the extension of prediction intervals, but also with the enhancement of actual wind speed. In addition, terrain effect is taken into consideration to study the relationship between forecast precision and geographic altitude, especially in Qinghai-Tibet Plateau.

P MISC-7: Evaluating the performance of RBMP-based multi-model ensemble forecasts for 2 meter temperature

AUTHORS: Zhu, Yuejia; Guan, Hong; Cui, Bo; Zhu, Yuejian; Meng, Wen.

Presenting author e-mail: zhuyuejia_124@163.com; Affiliation (all authors): Anhui Meteorological Bureau/ CMA, Hefei, China

ABSTRACT: Many studies have demonstrated the advantages of multi-model ensembles. In our study, we access the performance of multi-model ensemble forecasts using common metrics. Those metrics will include the investigation of spread-skill relationship and quantitative scores such as continuous ranked probability score (CRPS) and Relative Operating Characteristic (ROC). We generate multi-model ensembles by Recursive Bayesian Model Process (RBMP), which is unequally blending three global ensemble forecast systems from the NCEP, CMC and FNMOC, mainly based on the Bayesian Model Averaging (BMA) method introduced by Raftery et al. (2005) and the station-based BMA codes created in MDL, while the decaying averaging method is also applied in parameter estimation and 2nd moment adjustment. To conduct multi-model ensembles, the global grid-based RBMP for 2 meter temperature is initialized on 1 September 2013, and the evaluation period begins on 1 December 2013 and extends through 30 November 2014. The performance of RBMP is verified using both ERA-interim grid analysis and Continental United States (CONUS) observations. In terms of common verification scores, RBMP improves ensemble forecast skill at all lead times for all seasons, especially for short lead times in winter.

P MISC-8: Estimation of effectiveness of thunderstorms observation and their forecast by the instability indices

AUTHORS: Shpyg, Vitalii; Budak, Igor.

Presenting author e-mail: vitold82@i.ua; Affiliation (all authors): Ukrainian Hydrometeorological Institute, Kyiv, Ukraine

ABSTRACT: Dangerous weather phenomenon such as thunderstorm, which is related with zones of active convection, refers to mesoscale. It forms in results of difficult interaction of atmospheric processes of synoptic scale and underlying surface. Horizontal scale of convective cells is less in two-three times and more than distance between surface meteorological stations of observation network. And comparatively small is duration of thunderstorms and heavy rainfalls. Actually, last two circumstances define difficulties of their registration, investigation and forecasting. Problems investigation and forecasting depend on reliability of registration. Therefore first of all it was necessary to estimate of effectiveness of thunderstorms observation by surface network. For the data processing and further analysis, the data Doppler meteorological radar of International airport "Boryspil", which is situated in Kyiv region on the north of Ukraine, and data of surface meteorological stations for period, covered 2013-2014 years. Radar archive consists from range of decoded images with information about spatial distribution of different meteorological phenomena and cloudiness altitude in zone of radar operation. For the diagnostics of the possibility and the fact of phenomenon by characteristics of radar reflectivity (reflectivity, radial velocity, angle of place and etc.) the set of empirical functions, which are realized in software of the radar, were used. For both data types the software was developed on PHP, C and R languages. It was done for decoding and recording radar and surface observations to the data bases on SQLite with goal of automation of the processing. Different instability indices have been used for investigations of separate cases and in operational forecast. In this research Total Totals Index, SWEAT Index, Showalter Index, KI and LI were calculated by output data of the WRF ARW and WRF NMM models. Hereinafter they were compared with data of the surface observations. In results it was obtained that existence meteorological stations network can observed near 30 % of thunderstorm events. That is, fixation of the coming of events doesn't reach their numbers observed by radar or, in other words, network understates their quantity significantly. Also it was shown that using of instability indices steadily lead to false alarms at the expense of existence of area of uncertainty in distributions of values of index for the events with and without thunderstorm. Therewith their distributions depend weakly from type of the NWP model and its spatial discretization.

P MISC-9: Evaluation of the impact of vertical resolution on the numerical weather prediction

AUTHORS: Lee, Eun-Hee; Lee, Eunjeong; Choi, In-Jin.

Presenting author e-mail: eh.lee@kiaps.org; Affiliation (all authors): Korea Institute of Atmospheric Prediction Systems, Seoul, South Korea

ABSTRACT: The impact of changes (L28 → L100) in vertical resolution on numerical weather prediction (NWP) is studied using the Global/Regional Integrated Model System-Global Model Program (GRIMs-GMP) with the physics package developed by the Korea Institute of Atmospheric Prediction Systems. In a series of medium-range forecast experiment during July 2013 and Feb 2014, the sensitivity of atmospheric structure is examined for various spatial resolutions. At higher vertical resolution (L100) simulates colder troposphere with increasing resolved wind field, while it produces warmer stratosphere. As a result, increased temperature bias appears most prominently at mid-latitudes compared with simulation with the lower vertical resolution. Consistent with these temperature changes, cloud amount increases with vertical resolution increase, showing enhanced weak precipitation and reduced heavy precipitation. It results overall decrease of precipitation prediction skill, which suggest that more resolved cloud and precipitation in microphysics may affect to precipitation distribution in NWP.

P MISC-10: 17 years "Berlin Weather Forecasting Contest" - a unique verification tool

AUTHORS: Knüpffer, Klaus; Hoffmann, Jan.

Presenting author e-mail: klaus@mswr.de; Affiliation (all authors): METEO SERVICE weather research GmbH, Berlin, Germany

ABSTRACT: Weather forecasting contests are a unique and efficient tool to prove the skills of forecasters and forecasting algorithms. The results provide real-time comparative verification response immediately after the observations are present. The Berlin Weather Forecasting Contest (www.wetterturnier.de) is running for more than 17 years now: Each Friday more than 100 participants - among them more than 10 different MOS forecasts (from different MOS developers, based on different models) - take part with their forecasts for 12 weather elements for Saturday and Sunday for 5 places in Germany, Austria and Switzerland. Verification of the results will be presented. This allows interesting insights regarding - the performance of different numerical models as basis for MOS, - the performance of MOS of different providers, - the positive effect of blending of two or more Models in a MOS-Mix, - the performance of human and machine forecasts and - the general improvement of forecast accuracy over the years. MOS-Mix forecasts are typically ahead of the single-model MOS forecasts, and only a few forecasters are able to score better than the best MOS-Mix forecast which is available to them a few hours before deadline (Friday 15 UTC). The general progress in short range weather forecasting per decade is "Look one day more ahead with the same accuracy as 10 years ago". This can also be found in the tournament data. Participation is free for anyone who wants to sharpen his forecast skills or to evaluate new forecasting algorithms.

P MISC-11: Synoptic verification characteristics of operational GRAPES-GFS model forecast

AUTHORS: Yu, Gong; Kan, Dai; Shunan, Yang; Jian, Tang; Fang, Zhang; Ning, Hu.

Presenting author e-mail: heygy@live.cn; Affiliation (all authors): National Meteorological Center of China Meteorological Administration, Beijing, China

ABSTRACT: As the GRAPES-GFS model has been put into operational work in early 2016, the NMC (National Meteorological Center of China Meteorological Administration) also deployed operational synoptic verification on GRAPES-GFS correspondingly. After the verifications on the whole 2016 year real time forecast and parts of 2013-2015 back calculate data, there were some results had been excavated. Therefore, to help improving the GRAPES-GFS developments and application, this work gathered all the results we had got, and summarized up into several systematic characteristics. There are 38 heavy rainfalls and 8 hot weather verified. Start from the forecast quality of precipitation and hot weather, Synaptic weather systems and atmospheric physical factors were checked to find the direct causes of the precipitation biases or advantages. The results show that, there are some advances been made in short-range precipitation forecast, but still shows north bias in some convective rainband forecast. Precipitation forecasts are weaker than observation in some weak synoptic system background. Wet bias northern to the rainband and stronger bias of subtropical vortex were also found in some cases.

P MISC-12: Verification of Open-Ocean Environments of Wind Speed and Wave Heights

AUTHORS: Lee, Raymund C.; Wittmann, Paul.

Presenting author e-mail: raymond.c.lee1@navy.mil; Affiliation (all authors): Fleet Numerical Meteorology and Oceanography Center, Monterey, United States of America

ABSTRACT: Fleet Numerical Meteorology and Oceanography Center (FNMOC), under the direction of the Commander, Naval Meteorology and Oceanography Command, has the mission to operate global and limited area meteorological and oceanographic numerical weather prediction (NWP) models in support of the US DoD in operations, exercises, humanitarian assistance and disaster relief. Verification of ocean surface winds and waves is made more difficult due to a lack of conventional METAR or SYNOP observations typically found on land. Satellite based observations of wave height and surface wind speed/direction have been leveraged to provide wind and wave verification for our NWP models. Acknowledging the slight operational retrieval error of these observations we use scatterometer and altimeter retrievals in an automated near-real-time verification of the global and limited area regions over ocean environments to provide guidance for the fleet forecasters and model developers.

Annalen der Meteorologie

Nr. 50 Der globale Klimawandel - im Spiegel der wissenschaftlichen Veröffentlichungen : Meilensteine der Publikationsgeschichte zum anthropogenen Treibhauseffekt seit 1824; Ausstellungskatalog 30.01. - 28.04.2017 in der Deutschen Meteorologischen Bibliothek. 2017.

Nr. 49 10. Annaberger Klimatage 2016 : 11. - 12. Mai 2016 in Annaberg-Buchholz; Tagungsband mit Beiträgen anlässlich der Gründung der Wetterwarte Fichtelberg vor 100 Jahren. 2016

Nr. 48 Phenology 2015: Third International Conference on Phenology, 5.-8. October 2015, Kusadasi (Turkey), Programme and Abstracts. 2015

Nr. 47 METTOOLS IX: Programm, Übersicht, Kurzbeiträge. Tagung des Fachausschusses Umweltmeteorologie FAU MET der Deutschen Meteorologischen Gesellschaft DMG e.V. vom 17. bis 19. März 2015 in Offenbach (Main). 2015

Nr. 46 Goethes weiteres Erbe : 200 Jahre Klimastation Jena (Beiträge des Jubiläumskolloquiums „200 Jahre Klimamessstation Jena“. 2013

Nr. 45 125 Jahre Wetterstation Geisenheim 1884 - 2009 (Jubiläumsband mit Beiträgen aus der agrarmeteorologischen Forschung und Beratung). 2009

Nr. 44 Programme & Extended Abstracts, 30th International Conference on Alpine Meteorology : 11 - 15 May 2009, Badnerhalle, Rastatt, Germany. 2009

Nr. 43 125 Jahre Deutsche Meteorologische Gesellschaft : Festveranstaltung am 7. November 2008 in Hamburg. 2008

Nr. 42 50 Jahre Überwachung der Radioaktivität in der Atmosphäre durch den Deutschen Wetterdienst : Festveranstaltung am 8. - 9. Juni 2005, Langen, Deutschland. 2006

Nr. 41 17th International Congress of Biometeorology - ICB 2005. 2005

Nr. 40 SIRWEC 2004 - 12th International Road Weather Conference, Bingen. 2004, CD-ROM

Nr. 39 4. Fachtagung BIOMET (19. - 20.April 1999 in München) des Fachausschusses BIOMET der DMG gemeinsam mit der Gesellschaft zur Förderung Medizin-Meteorologischer Forschung e. V. 1999

Nr. 38 Herbstschule Radarmeteorologie 1998 vom 2. bis 6. November in Oberpfaffenhofen. 1998

Nr. 37 Deutsche Meteorologen-Tagung vom 14. bis 18 September 1998 in Leipzig, Band 1 und 2. 1998

Gesamtliste der „Annalen der Meteorologie“ unter www.dwd.de/bibliothek.

Deutscher Wetterdienst
Fachinformationsstelle und Deutsche Meteorologische Bibliothek
Frankfurter Straße 135
63067 Offenbach