Forecast Verification in a Polar Framework

Pertti Nurmi
Finnish Meteorological Institute

Member of WWRP Polar Prediction Project (PPP) Steering Group
Member of WWRP Joint Working Group on Forecast Verification Research (JWGFVR)

Acknowledgements:
- JWGFVR
- ECMWF
- M. Matsueda

Photo © Timo Vihma
A few words about JWGFVR: Goals & activities

Verification component of WMO WWRP, in collaboration with WGNE, WCRP, CBS

- Serve as a focal point to develop and promote new verification methods
- Promote importance of verification (as vital part of experiments)
- Promote collaboration among verification scientists, model developers, forecast providers AND end-users (customers)
- Emphasize user-aspects of forecast verification ↔ Impacts
- Provide training on verification methodologies
  ✓ 3 extensive tutorials organized so far; Next in spring 2014?
- Does NOT provide “verification services” per se …
References ...

<table>
<thead>
<tr>
<th>WWRP 2009 - 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models</td>
</tr>
<tr>
<td>Revision 2, October 2008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WWRP 2012 - 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommended Methods for Evaluating Cloud and Related Parameters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>To appear in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verification of tropical cyclone forecasts</td>
</tr>
</tbody>
</table>

8 papers by JWGFV members in Special Issue (June 2013) of *Met. Apps.*
- incl. a "lead paper" Progress and challenges in forecast verification

References: Only few addressing Polar Verification per se

Based on survey to find relevant papers covering past c. 10 years:

8. Lammert et al., 2010. Comparison of three weather prediction models with buoy and aircraft measurements under cyclone conditions in Fram Strait. Tellus Ser. A, 351-376.
**Service-oriented Research**

- **Societal and Economic Research Applications (SERA)**
- **Verification**
  - Modeling perspective \( \Leftrightarrow \) User perspective !

**Underpinning Research**

- Predictability and Diagnostics
- Teleconnections

**Forecasting System Research**

- Observations
- Data Assimilation
- Modelling
- Ensemble Forecasting
User perspective!

Road surface friction forecast verification on northern roads

RWS "Anjala"

RWS "Utti"

User perspective!
New Issue: “Seamless” or consistent verification across all scales ⇔ applying same verification measures to all forecasts, to allow comparison.

PPP verification activities

Wide range of scales, models, methods, variables, observations …

… NOT trivial to define a general verification setup to cover everything, systematically …
A. Observations

✓ Identification, definition & establishment of optimal, high-resolution observing networks
✓ Utilization of in-situ & remote sensing observations
✓ “Invention” of new, mobile (?) observing means; cf. road transport
✓ Issues with complex terrain + surface properties

B. Raise awareness of the necessity for comprehensive verification

C. Verification methods and metrics

✓ Dedicated metrics for dedicated, high-impact polar phenomena
  • Low cloud, fog, visibility, blizzards, wind, temperature extremes
✓ Verification methods R&D
  • Exploration of existing vs. new, upcoming verification metrics
✓ Definition and adoption of “seamless verification” to cope with seamless forecasting
✓ Address both deterministic and probabilistic forecasts (obviously)
GA1. Review & examine present verification state-of-the-art
✓ Literature review
✓ Applicability to polar specific phenomena and applications
✓ All forecast variables and types & all forecast scales: hourly-to-seasonal
✓ Seamless applicability, multi-dimensionality

GA2. Distinguish key user-relevant, high-impact weather elements (not forgetting sea ice)
✓ Low cloud, fog, visibility, blizzards, wind, temperature extremes
✓ Definition of variables and their temporal and spatial scales, followed by verification specifications for each

GA3. Try to devise and apply polar-tailored – potentially new – verification techniques
✓ User-relevance ⇔ Impacts

Loosely following Implementation Plan
GA4. Carry out polar vs. mid-latitude verification comparison

- Verification of existing forecasting systems \(\iff\) Comparison of past and present forecast performance and progress
- Compare polar vs. non-polar (mid-latitude) forecast performance
- Systematic comparison between different Forecast Centres
- Investigation of polar lows
  - Possibly utilize methodology like for tropical cyclones

GA5. Is there potential / interest to develop spatial verification techniques for polar areas?

- Feasibility with lack of data? Only polar orbiter data available? Only for cloud verification? Can we distinguish cloud from ice? ...
- Needs motivation and commitment \(\iff\) Potential collaboration with spatial forecast verification methods inter-comparison initiatives and programs

Loosely following Implementation Plan
GA6. Define and adopt “headline” performance measures

- To monitor polar fc performance throughout the 10-year project lifetime
- Comparison between different forecasting systems and Centres

GA7. Devise and perform user-oriented verification

- Distinguish specific (end-) users and their requirements
- Define & apply “simplified” verification metrics addressing end-users
- Provide guidance to Weather Services to adopt and apply meaningful user-oriented verification measures
- Forecast value (c/b; C/L) issues addressing impacts ⇔ SERA

GA8. Analyze present and explore new observation means

⇔ YOPP observation & verification strategy

- E.g. mobile observation platforms; utilization of non-conventional data; new telecommunication techniques facilitating rapid applicability
- Observation uncertainties

Loosely following Implementation Plan
GA9. YOPP - Polar test bed(s)
- Enhanced verification utilizing comprehensive “Verification Toolboxes”
- Potentially build up a Real-Time Forecast Verification System (RTFVS)
- YOPP impact studies and post-YOPP consolidation

GA10. Identify data needs, organize data collection, storage and access ⇔ YOPP data centre (ref. TIGGE)
- Common data formats & platforms to ease access and encourage use

GA11. Set up & launch a centralized verification effort
- Many Centres, possibly, apply own differing non-uniform metrics
- Seek for potentially interested host Meteorological Service(s)

GA12. Set up a dedicated verification expert team
- PPP expert team members enforced by verification “enthusiasts”
- Lead Centres of verification; WMO meso-scale working group etc…

Loosely following Implementation Plan
Desirable specific properties for a verification measure:

- Dependency on the verification, or analysis, grid should be minimised.
- Dependency on spatial and temporal scales and sampling of observation data should be minimised.
- Behaviour should not depend on the base value, i.e. on the magnitude of verified quantity.
- Behaviour should not depend on the base rate, i.e. climatology.
- Should remain useful for rare events: Most conventional measures become unusable beyond c. 90 percentile.
- Should converge quickly for relatively small samples.
- Should be accompanied by estimates of uncertainty - error bars.
- Should take both hits and false alarms into account, for categorical fcs.
- Should be “proper”, “equitable” and not reward “hedging”

⇒ No currently available metrics satisfy all these !!!
Examples of some relatively new verification metrics / methods

"Traditional" scores tend to zero with the rarity of the event, i.e. are highly dependent on base rate (i.e. local climatology)!

Looking at multiple scores at one time
Only need to plot POD and 1-FAR
(Success Ratio)

(From Roberts et al. 2011; after Roebber 2009 and C. Wilson 2008)
Lots of activity during past – and probably during coming several years

Designed to diagnose spatial structures like precipitation areas, fronts …
⇔ Cover different scales

Provide information on error in physical terms

Account for uncertainties in location and timing

Typically utilize remote sensing satellite and/or radar data

Would require a high density observation network!

Neighborhood methods, Fractions Skill Score, Feature-based methods, CRA, SAL, MODE, etc…

Starting to penetrate to ensemble forecast verification
Evolution of ECMWF scores comparison northern and southern hemispheres

... but how about polar prediction forecast quality?
Daily scores for TIGGE

- Running mean:
  - 365-day
  - 91-day
  - 31-day

Areas:
- Hemispheres
- Polar regions
- Midlatitudes

Scores:
- Wrt own analysis:
  - ACC (control run)
  - ACC (ensemble mean)
  - RMSE (control run)
  - RMSE (ensemble mean)
- Wrt ERA-Interim:
  - ACC (control run)
  - ACC (ensemble mean)
  - RMSE (control run)
  - RMSE (ensemble mean)
- RPSS

Variables:
- Z500
- T850
- T2m
- U850
- V850
- U200
- V200

Figure: Skill comparison of TIGGE medium-range ensemble forecasts

ACC Z500 control run (OCT2006–MAR2013)

- Northern mid-latitude (20°–60°N)
- Southern mid-latitude (20°–60°S)

- 365-day running mean, wrt own analysis

Days 3, 5, 7, 9, 15

Go to scores for NCEP GEFS
Go to the main page

http://tparc.mri-jma.go.jp/TIGGE/index.html © Dr. Mio Matsueda

pertti.nurmi@fmi.fi
Daily scores for TIGGE

- Running mean:
  - 365-day
  - 91-day
  - 31-day

Areas:
- Hemispheres
- Polar regions
- Midlatitudes

Scores:
- WRT own analysis:
  - ACC
    - Control run
  - ACC
    - Ensemble mean
  - RMSE
    - Control run
  - RMSE
    - Ensemble mean

WRT ERA-Interim:
- ACC
  - Control run
  - ACC
  - Ensemble mean
  - RMSE
  - RMSE
  - Ensemble mean
  - RPSS

Variables:
- Z500
- T850
- T2m
- U850
- V850
- U200
- V200

Go to scores for NCEP GEFS
Go to the main page

http://tparc.mri-jma.go.jp/TIGGE/index.html © Dr. Mio Matsueda

Skill comparison of TIGGE medium-range ensemble forecasts
ACC Z500 control run (OCT2006–MAR2013)

365-day running mean, WRT own analysis

Day 3
Day 5
Day 7
Day 9
Day 15

Anomaly correlation coefficient
Comparison of TIGGE medium-range ensemble forecasts (Z500) +168hr

Anomaly Correlation

Northern mid-latitude, 2012.12

Arctic, 2012.12

© Dr. Mio Matsueda

pertti.nurmi@fmi.fi

http://tparc.mri-jma.go.jp/TIGGE/index.html

© Dr. Mio Matsueda
Comparison with other centres (2000-2012)

Arctic RMS Error

Antarctic RMS Error

Arctic ACC

Antarctic ACC

Note diff. y-axis scale

Much more ECMWF diagnostics in next talk
Obs availability, spring 2013

Mean Sea Level Pressure (Pa): Surface Obs

- Northern Polar Region (CBS area 90N-60N)
- Southern Polar Region (CBS area 60S-90S)

(Acknowledgement: Marion Mittermaier)
North - South Comparison

RMSE, spring 2013

Mean Sea Level Pressure (Pa): Surface Obs

- UK-GM
- EC-GM
- US-GM

Northern Polar Region (CBS area 90N-60N)
Southern Polar Region (CBS area 60S-90S)

NB: UK @ full resolution; EC & US @ coarser CBS grip resolution

(Acknowledgement: Marion Mittermaier)
North - South Comparison

RMSE, spring 2013

- UK-GM
- EC-GM
- US-GM

Mean Sea Level Pressure (Pa): Surface Obs
Equalized and Mean ed from 7/3/2013 00Z to 4/6/2013 12Z

NB: UK @ full resolution; EC & US @ coarser CBS grip resolution

Acknowledgement: Marion Mittermaier
✓ No such thing as observed “truth”
  ▪ Regardless how good your observations, they are always estimates!
  ▪ Forecast verification would require knowing the “truth”, however

✓ Observational uncertainty need to be taken into account
  ▪ E.g., how well do nearby observations match each other?
  ▪ Quality checking of observations
    o Removal of gross errors, instrument and reporting errors; biases

✓ Observations generally are “more true” than model analyses

  ⇔ Utmost care if using model analysis as verifying “truth”

  ⇔ Analyses typically are highly model dependent!
    ✓ Especially so in polar regions with lack of observations

  ⇔ Analyses suited for comparison between versions of same model -
  e.g. operational vs. experimental suite – rather than comparing different
  models against each other
Observations are THE cornerstone of forecast verification!

You always get best verification scores when using your own analysis:

- Own model climatology brings advantage
- Differences largest in the tropics and at low levels

Repeat this kind of experiment for the Polar regions.
Predictability - Free atmosphere

✓ 1990 ⇔ 4 days
✓ 2000 ⇔ 5 days
✓ 2010 ⇔ 6 days ⇔ Expected increase ⇔ 1 day / decade

Predictability ⇔ ECMWF “headline” measure

ACC of Z 500 remains above 80 %

Need to set targets for polar prediction performance!
Predictability - Surf. weather ➔ End-user perspective

- 1995 ➔ 2 days
- 2010 ➔ 3.5 days ➔ Expected increase ➔ 1 day / decade

“1 – SEEPS” of 24 hr precipitation remains above 45 %

Need to set targets for polar prediction performance!
1. Investigate and test present and new, upcoming verification measures
2. Utilize verification as a means to assist observing system design - YOPP
3. Agree on (at some stage) a common set of verification metrics (for YOPP)
4. YOPP data centre ⇆ Include a verification module
5. Seamless forecasting calls for seamless verification
6. Focus on forecasting capabilities of meaningful high-impact weather events, taking into account (end-) user aspect ⇆ Impacts ⇆ SERA
7. Potentially set up a real-time verification framework/system
8. Set up a verification expert team ⇆ linkages + outreach and education
   ⇆ Verification is MUCH more than bias, RMSE & ACC...

"Summary"

Interest in polar region dedicated verification has clearly increased since the initiation of PPP!
thanks You