WWRP/THORPEX Workshop on Polar Prediction, ECMWF, 24-27 June 2013

Forecast Verification in a Polar $\frac{\text{(PPP)}}{\text{(YOPP)}}$ 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)

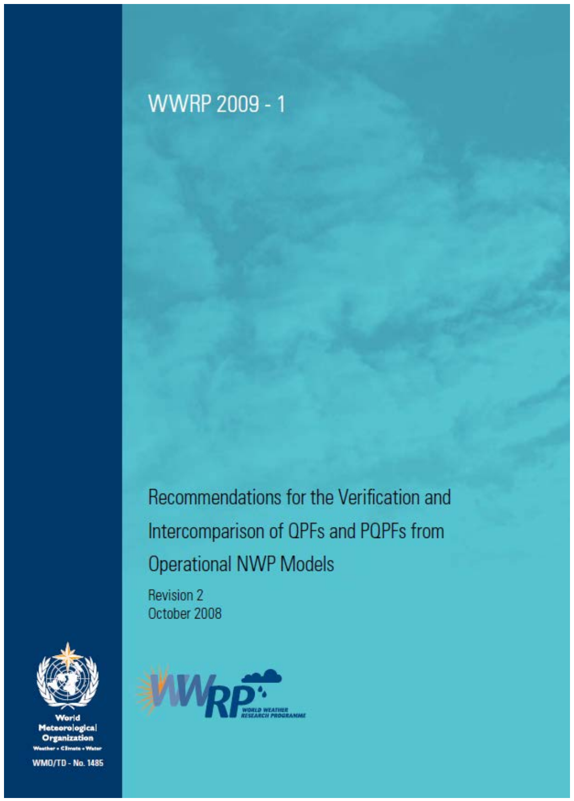
Acknowledgements:
- ✓ *JWGFVR*
- ✓ *ECMWF*
- ✓ *M. Matsueda*

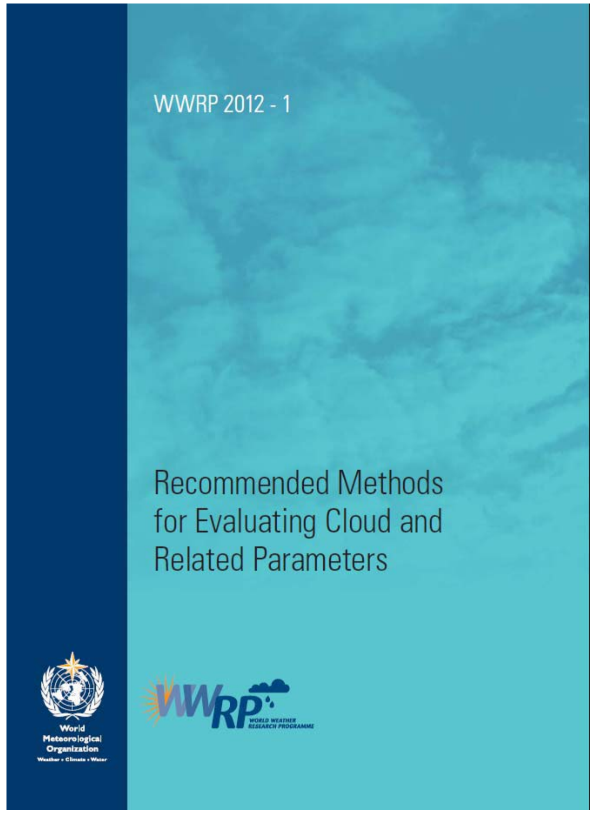
A few words about JWGFVR : Goals & activities

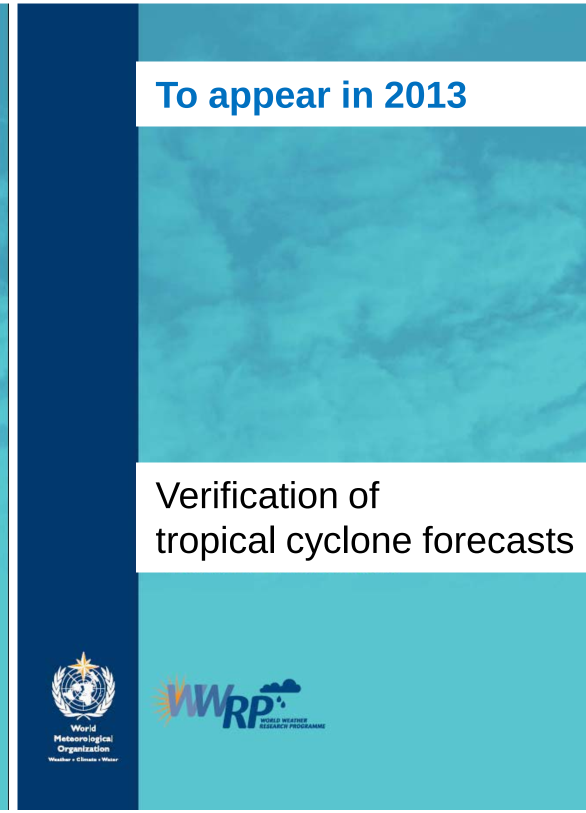
Verification component of WMO WWRP, in collaboration with WGNE, WCRP, CBS

- **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 ...







8 papers by JWGFVR members in Special Issue (June 2013) of *Met. Apps.*

- incl. a *"lead paper"* Progress and challenges in forecast verification

Our popular website : www.cawcr.gov.au/projects/verification

References : Only few addressing Polar Verification per se

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

1. Bromwich et al, 2013. A comprehensive evaluation of Polar WRF forecast performance in the Antarctic. J. Geophys. Res: Atmospheres, 274-292.
2. Jakobson et al, 2012. Validation of atmospheric reanalyses over the central Arctic Ocean. Geophys. Res. Lett, 39.
3. Kilpeläinen et al, 2012. Modelling the vertical structure of the atmospheric boundary layer over Arctic fjords in Svalbard. Q. J. R. Meteorol. Soc, 1867-1883.
4. Tastula et al, 2012. Evaluation of Polar WRF from the Modeling the Atmospheric Boundary Layer over Antarctic Sea Ice in Autumn and Winter, Mon. Weather Rev.
5. de Boer et al, 2012. A Characterization of the Present-Day Arctic Atmosphere in CCSM4. J. Climate, 2676-2695.
6. Hines et al, 2011. Development and Testing of Polar WRF. Part III: Arctic Land. J. Clim, 26-48.
7. Liston & Hiemstra, 2011. The Changing Cryosphere: Pan-Arctic Snow Trends (1979–2009). J. Clim, 5691-5712.
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, 361-376.
9. Dworak & Key, 2009. Twenty Years of Polar Winds from AVHRR: Validation and Comparison with ERA-40. J. Appl. Meteor. Climatol, , 24-40.
10. Walsh et al, 2009. Arctic Cloud Fraction and Radiative Fluxes in Atmospheric Reanalyses. J. Clim, 2316-2334.
11. Fogt & Bromwich, 2008. Atmospheric Moisture and Cloud Cover Characteristics Forecast by AMPS. Weather Forecast, 914–930.
12. Hines & Bromwich, 2008. Development and Testing of Polar Weather Research and Forecasting (WRF) Model. Part I: Greenland Ice Sheet Meteorology. Mon. Wea. Rev, 1971-1989.
13. Schweiger et al, 2008. Relationships between Arctic Sea Ice and Clouds during Autumn. J. Clim, 4799-4810.
14. Tjernström et al, 2008. How Well Do Regional Climate Models Reproduce Radiation and Clouds in the Arctic? J. Appl. Meteor. Climatol, 2405-2422.
15. Jung & Leutbecher, 2007. Performance of the ECMWF forecasting system in the Arctic during winter. Quart. J. Roy. Meteor. Soc, 2007, 1327-1340.
16. Nordeng et al, 2007. Improvement of weather forecasts in polar regions. WMO Bulletin 56(4) – October 2007.
17. Rinke et al, 2006. Evaluation of an ensemble of Arctic regional climate models: spatiotemporal fields during the SHEBA year. Climate Dynam, 459-472.
18. Bromwich et al, 2005. High-Resolution Regional Climate Simulations over Iceland Using Polar MM5. Mon. Wea. Rev.
19. Tjernström et al, 2005. Modelling the Arctic Boundary Layer: An Evaluation of Six Arcmip Regional-Scale Models using Data from the Sheba Project. Boundary-L. Meterol, 337-381.
20. Guo & Bromwich, 2003. Evaluation of Polar MM5 Simulations of Antarctic Atmospheric Circulation. Mon. Weather Rev, 384-411.
21. Gauthier & Falkingham, 2002. Long Range Ice Forecasting Techniques in the Canadian Arctic – Initial Verification. Proc. 16th IAHR Int'l Symposium on Ice, Dunedin, New Zealand.
22. Mailhot et al, 2002. Mesoscale simulation of surface fluxes and boundary layer clouds associated with a Beaufort Sea polynya. J. Geophys. Res, 8031.
23. Bromwich et al, 2001. Mesoscale Modeling of Katabatic Winds over Greenland with the Polar MM5. Mon. Wea. Rev, 2290-2309.
24. Braaten & Tucker. A Ceiling and Visibility Prediction System Suitable for Antarctic Flight Operation. https://ams.confex.com/ams/pdfpapers/20168.pdf.
25. Bromwich et al. Evaluation of Operational Weather Forecasts for Antarctica from Polar MM5-AMPS. www.mmm.ucar.edu/mm5/workshop/ws02/Bromwich.pdf.
26. Guo et al. Verification of Polar MM5 Simulations of Antarctic Atmospheric Circulation. https://ams.confex.com/ams/pdfpapers/20475.pdf.
27. Leatham et al. Veerifying Model Forecasts of Arctic Fronts in Advance of Winter Storms in the Southern Plains. www.caps.ou.edu/reu/reu10/papers/Leatham.pdf
28. Sechrist et al, 1989. Forecasters Handbook for the Arctic. 9.2. Verification of Numerical Model Forecasts. Naval Environmental Prediction Research Facility. Monterey, CA. Tech. Rpt.
29. The International Antarctic Weather Forecasting Handbook.
30. www.antarctica.ac.uk/met/momu/International_Antarctic_Weather_Forecasting_Handbook/update%20Antarctic%20NWP%20-%20USA.php

Ref. Implementation Plan

Service-oriented Research

Societal and Economic Research Applications (SERA)	Verification

Modeling perspective ⇔ User perspective !

Underpinning Research

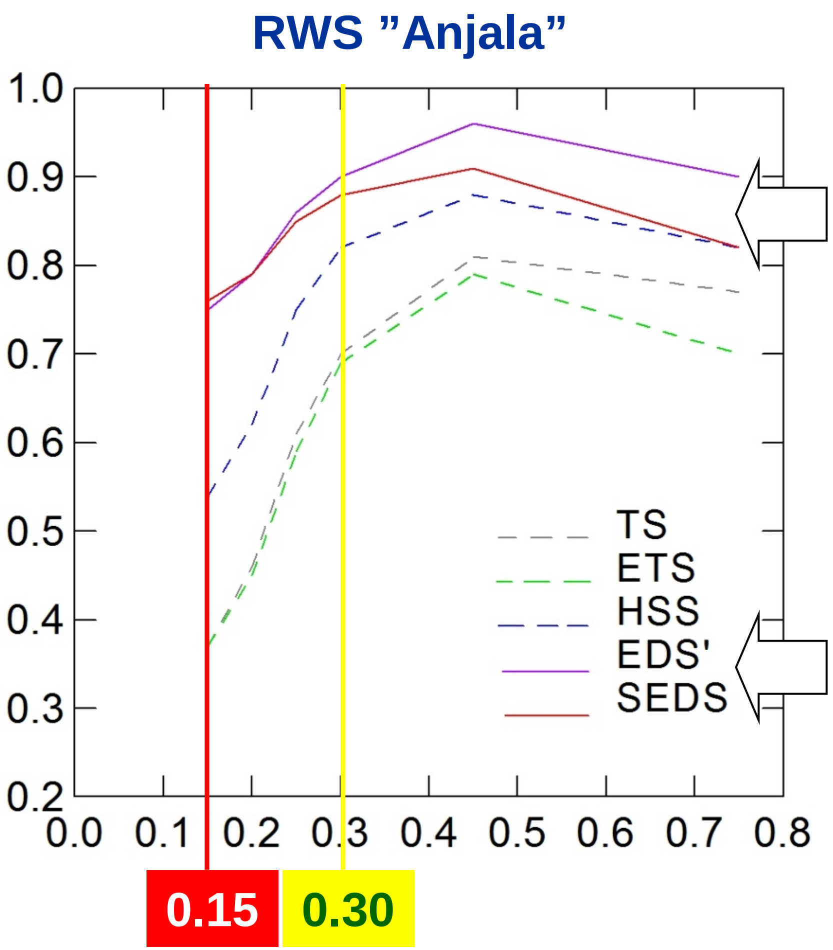
Predictability and Diagnostics	**Teleconnections**

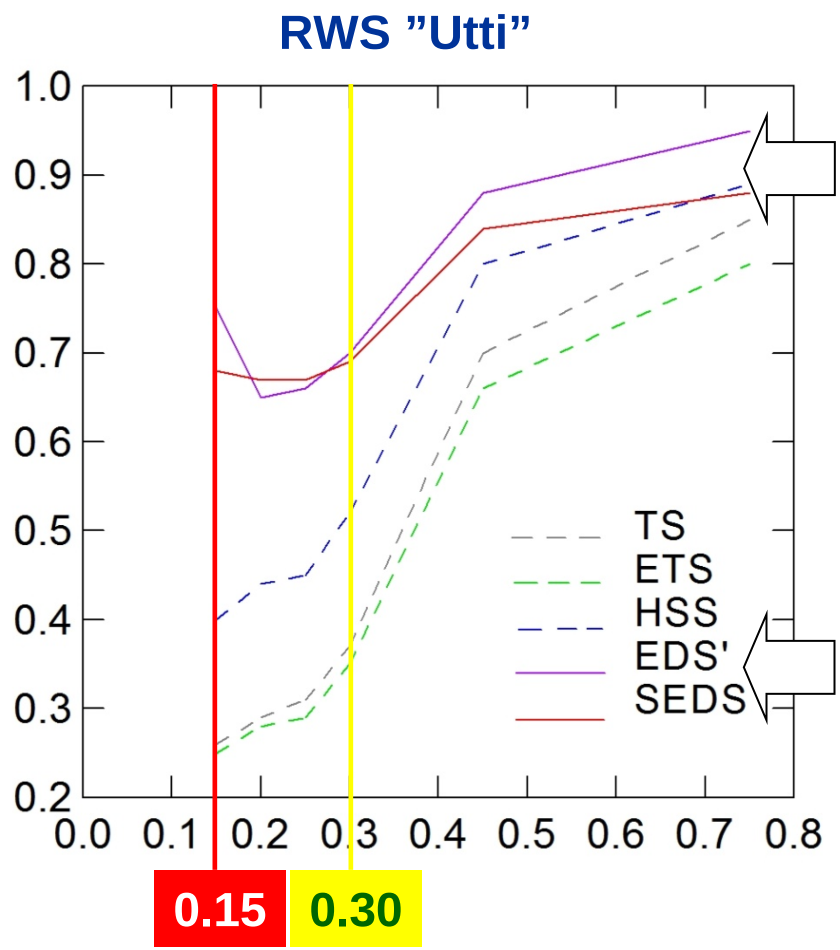
Forecasting System Research

Observations **Data Assimilation**	**Modelling** **Ensemble Forecasting**

User perspective !

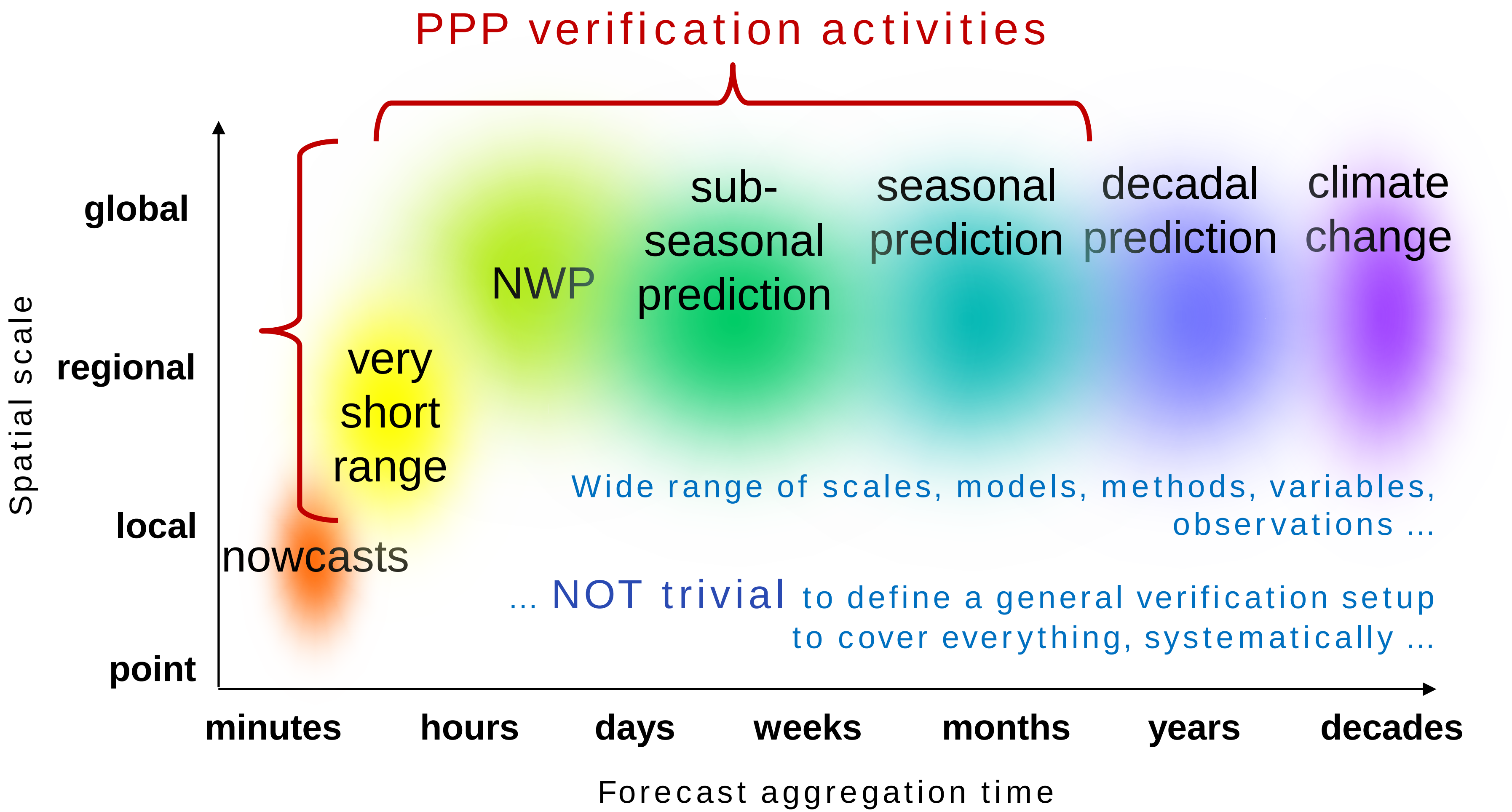
Road surface friction forecast verification on northern roads





Seamless verification ⇔ Seamless forecasting

New Issue : *"Seamless"* or *consistent* verification across all scales ⇔ applying same verification measures to all forecasts, to allow comparison



pertti.nurmi@fmi.fi

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)

... PPP Goals and Activities : i

Q1: 2013-15 Q3: 2017-19
Q2: 2015-17 Q4: 2019-21

GA 1. Review & examine present verification state-of-the-art — Q1

- ✓ Literature review
- ✓ Applicability to polar specific phenomena and applications
- ✓ All forecast variables and types & all forecast scales : hourly-to-seasonal
- ✓ Seamless applicability, multi-dimensionality

GA 2. Distinguish key user-relevant, high-impact weather elements **(not forgetting sea ice)** — Q1

- ✓ Low cloud, fog, visibility, blizzards, wind, temperature extremes
- ✓ Definition of variables and their temporal and spatial scales, followed by verification specifications for each

GA 3. Try to devise and apply polar-tailored – ***potentially new*** - verification techniques — Q1-3

- ✓ **User-relevance ⇔ Impacts**

Loosely following Implementation Plan

Preparation Phase 2012-2016 | **YOPP 2017-2018** | Consolidation Phase 2018-2022

Q1: 2013-15 Q3: 2017-19
Q2: 2015-17 Q4: 2019-21

GA 4. Carry out polar vs. mid-latitude verification comparison Q1

- ✓ Verification of existing forecasting systems ⇔ 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

GA 5. Is there potential / interest to develop ***spatial*** verification techniques for polar areas ? Q1-3

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

Loosely following Implementation Plan

... PPP Goals and Activities : iii

Q1: 2013-15 Q3: 2017-19
Q2: 2015-17 Q4: 2019-21

GA 6. Define and adopt ***"headline"*** performance measures

Q1-4

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

GA 7. Devise and perform *user-oriented* verification

Q1-4

- ✓ 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

GA 8. Analyze present and explore new observation means ⇔ YOPP observation & verification strategy

Q1-3

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

Loosely following Implementation Plan

Preparation Phase 2012-2016 | YOPP 2017-2018 | Consolidation Phase 2018-2022

... PPP Goals and Activities : iv

Q1: 2013-15 Q3: 2017-19
Q2: 2015-17 Q4: 2019-21

GA 9. YOPP - Polar test bed(s) — Q2-3

- ✓ Enhanced verification utilizing comprehensive "Verification Toolboxes"
- ✓ Potentially build up a Real-Time Forecast Verification System (RTFVS)
- ✓ YOPP impact studies and post-YOPP consolidation

GA 10. Identify data needs, organize data collection, storage and access ⬄ YOPP data centre (ref. TIGGE) — Q1-4

- ✓ Common data formats & platforms to ease access and encourage use

GA 11. Set up & launch a centralized verification effort — Q3-4

- ✓ Many Centres, possibly, apply own differing non-uniform metrics
- ✓ Seek for potentially interested host Meteorological Service(s)

GA 12. Set up a dedicated verification expert team — Q1-4

- ✓ PPP expert team members enforced by verification "enthusiasts"
- ✓ Lead Centres of verification; WMO meso-scale working group etc…

Loosely following Implementation Plan

Preparation Phase 2012-2016 → YOPP 2017-2018 → Consolidation Phase 2018-2022

Desirable specific <u>properties</u> for a verification measure :

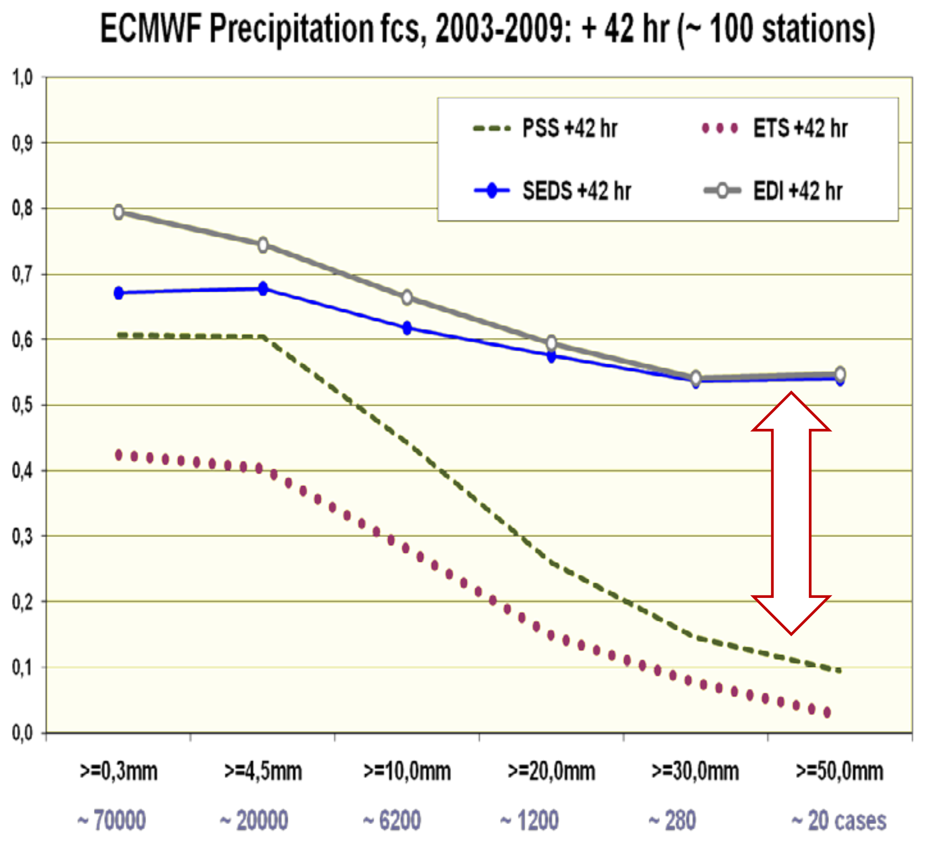
- ✓ **Dependency on the verification, or analysis, grid should be minimised**
- ✓ **Dependency on spatial and temporal scales and sampling of observation data should be minimised**
- ✓ **Behaviour should not depend on the base value, i.e.**
  on the magnitude of verified quantity
- ✓ **Behaviour should 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 !!!

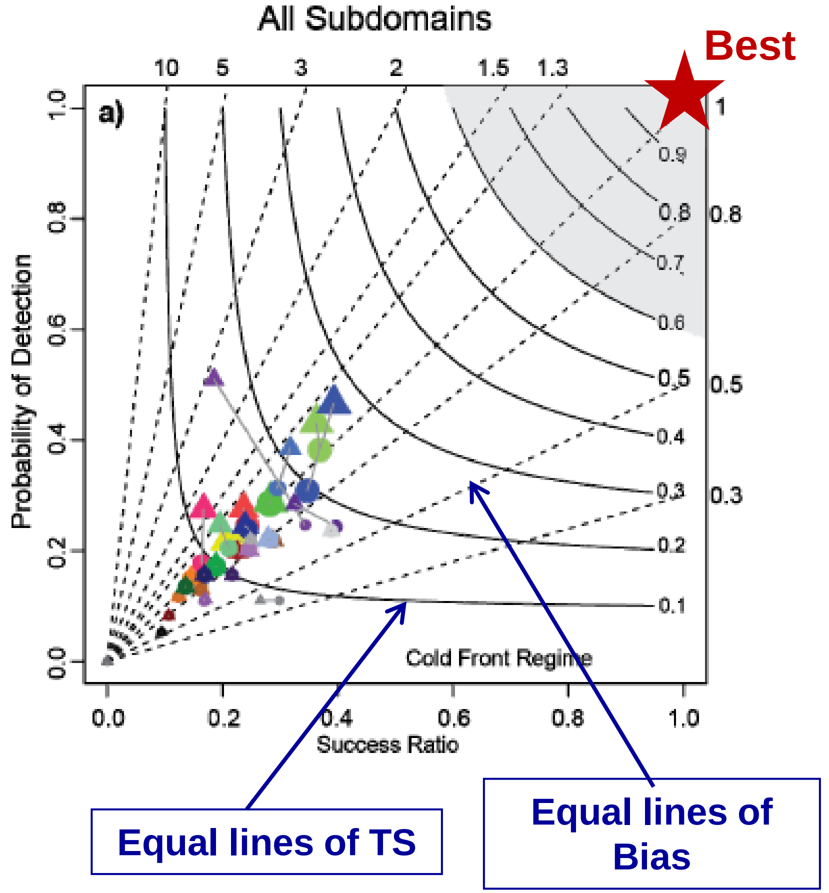
pertti.nurmi@fmi.fi

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)

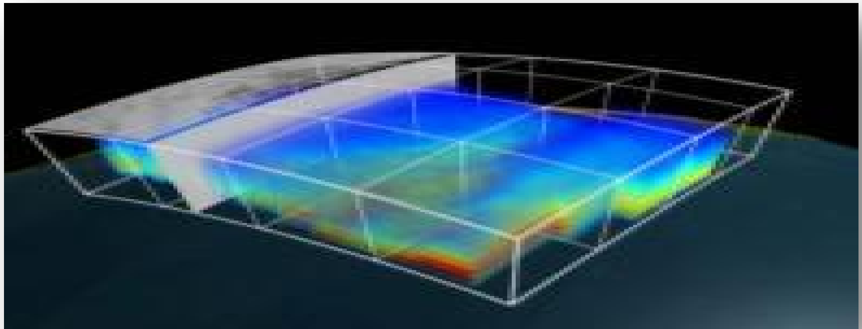


(From Roberts et al. 2011; after Roebber 2009 and C. Wilson 2008)

pertti.nurmi@fmi.fi

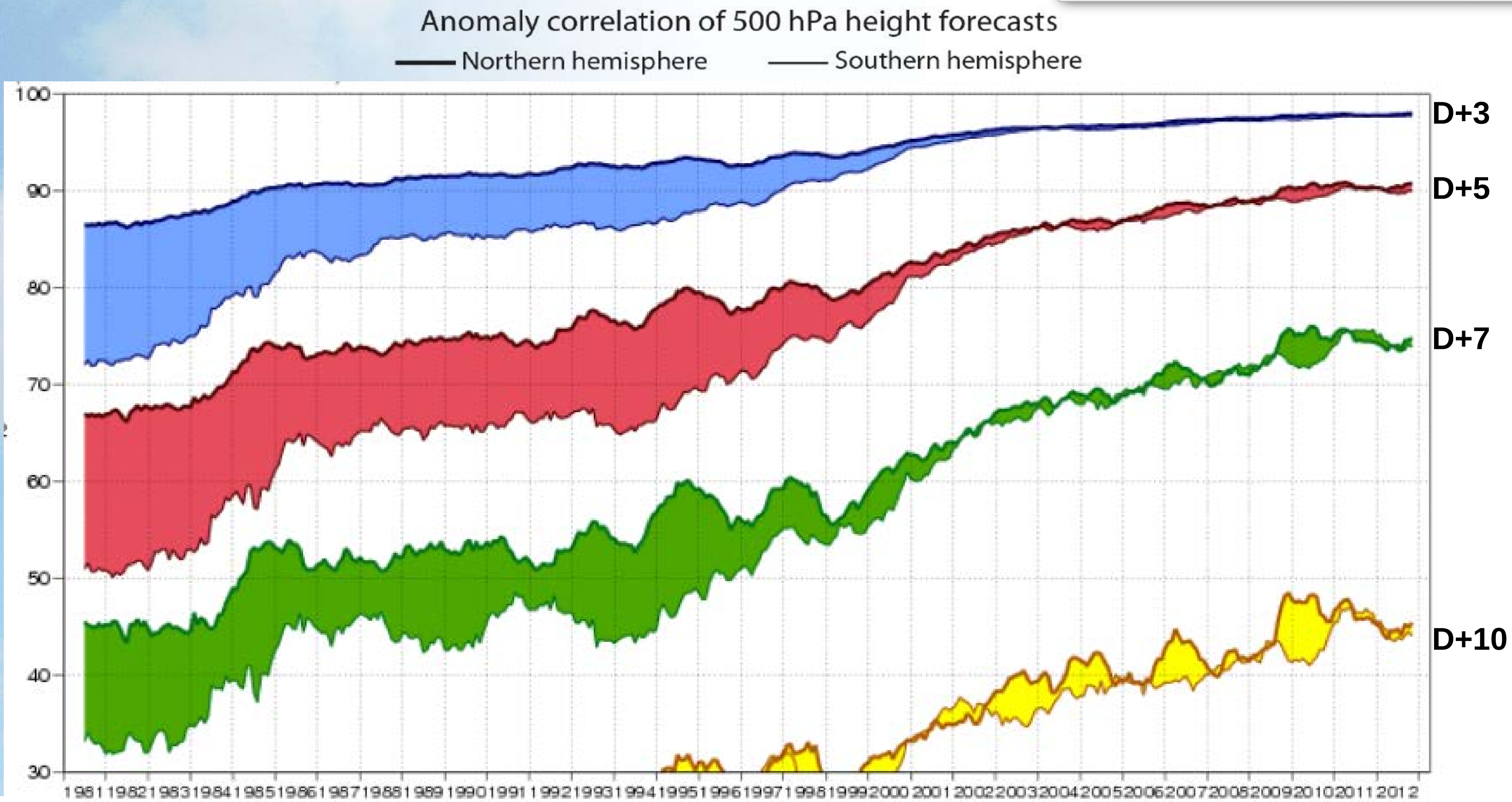
Spatial verification

- 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?



THORPEX
A World Weather Research Programme

GA-4, GA-11

Welcome to the TIGGE museum!

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

Daily scores for TIGGE

running mean:
- 365-day
- 91-day
- 31-day

Areas:
- Hemispheres
- Polar regions
- Midlatitudes

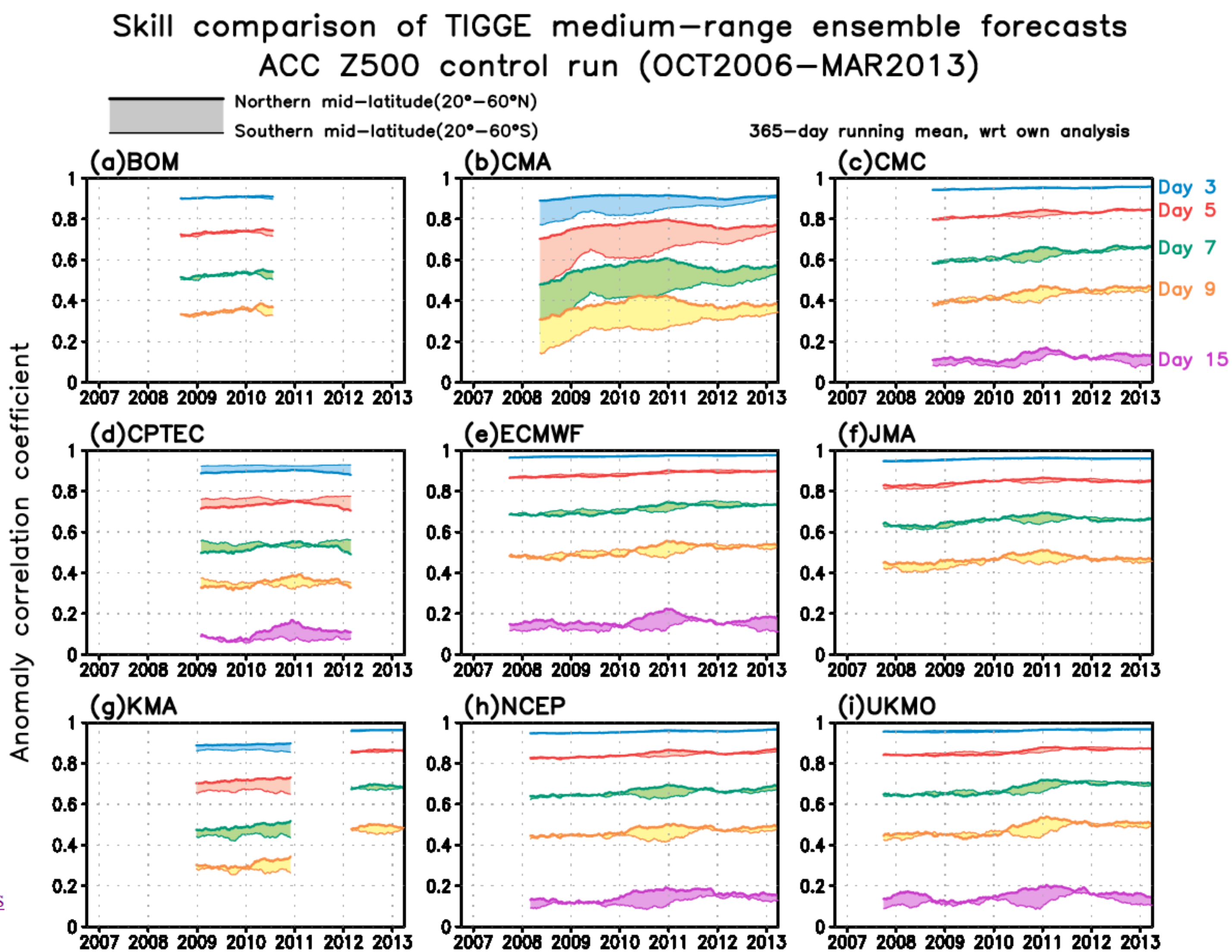
Scores:
wrt own analysis:
- ACC (control run)
- ACC (ensemble mean)
- RMSE (control run)
- RMSE (ensemble mean)

wrt ERA-Interim:
- ACC (control run)
- ACC (ensemble mean)
- RMSE (control run)
- RMSE (ensemble mean)
- RPSS

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

Go to scores for NCEP GEFS

Go to the main page



THORPEX
A World Weather Research Programme

GA-4, GA-11

Welcome to the TIGGE museum!

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

Daily scores for TIGGE

running mean:
- 365-day
- 91-day
- 31-day

Areas:
- Hemispheres
- Polar regions
- Midlatitudes

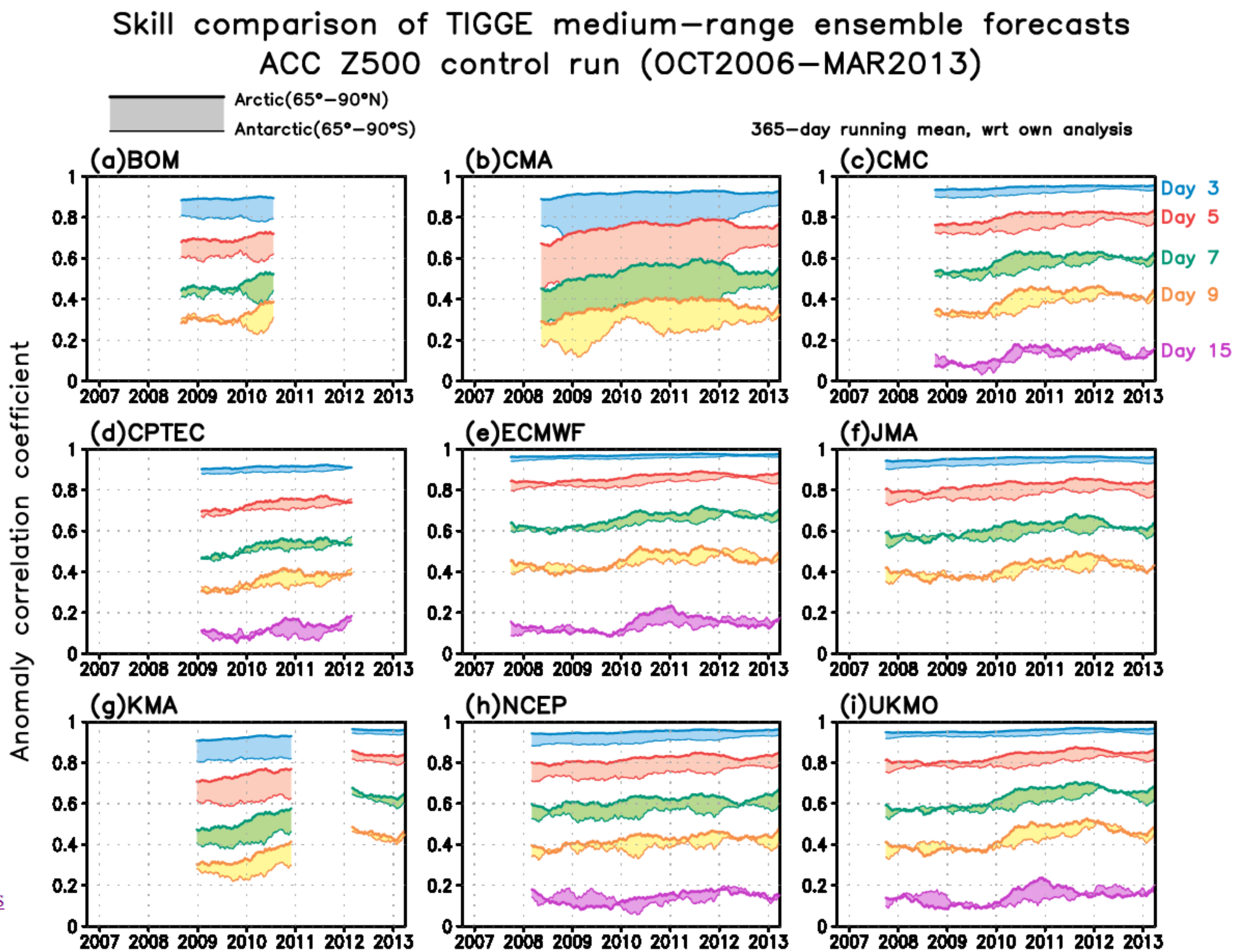
Scores:
wrt own analysis:
- ACC (control run)
- ACC (ensemble mean)
- RMSE (control run)
- RMSE (ensemble mean)

wrt ERA-Interim:
- ACC (control run)
- ACC (ensemble mean)
- RMSE (control run)
- RMSE (ensemble mean)
- RPSS

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

Go to scores for NCEP GEFS

Go to the main page



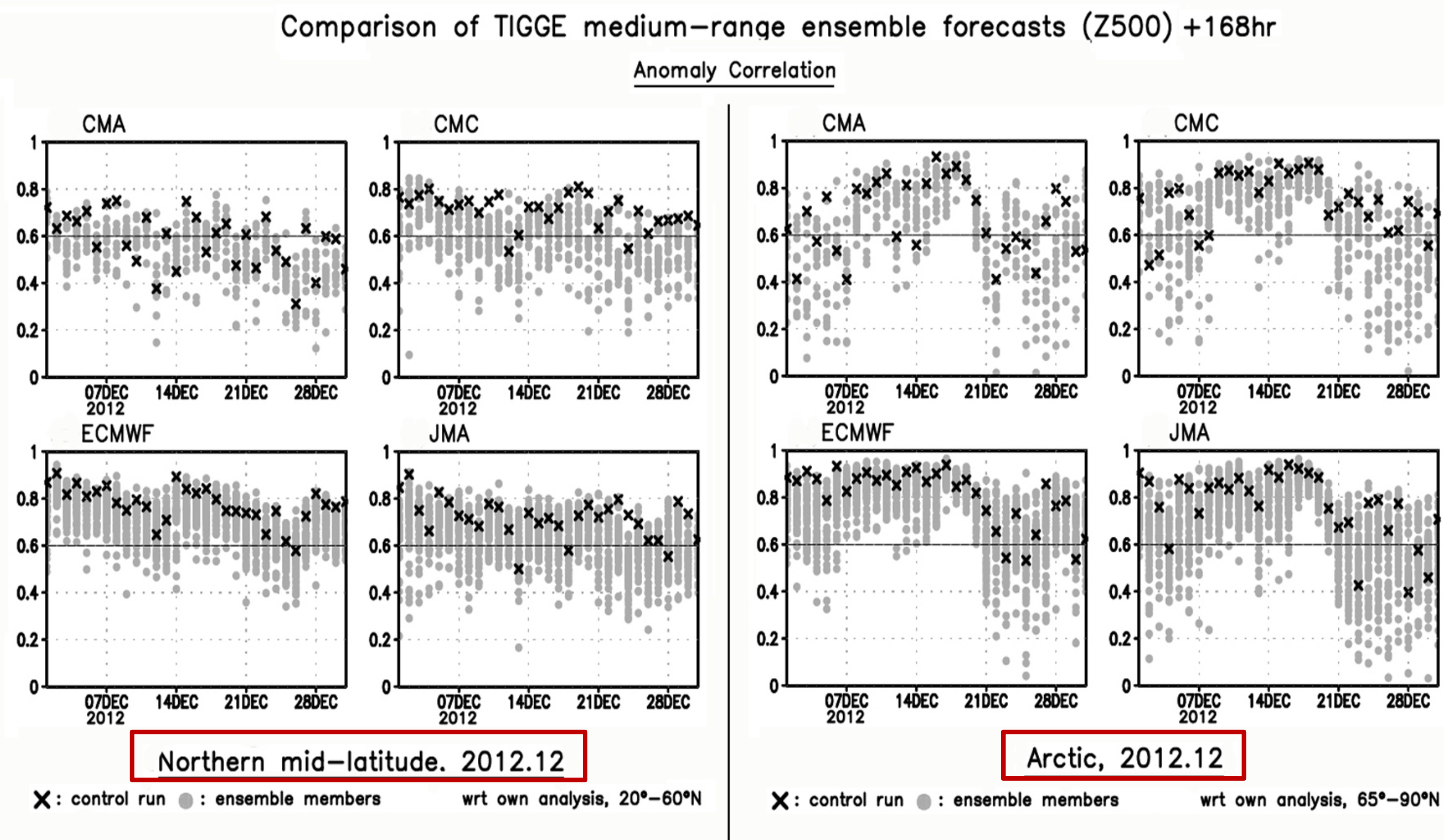
pertti.nurmi@fmi.fi

THORPEX
A World Weather Research Programme

Welcome to the TIGGE museum!

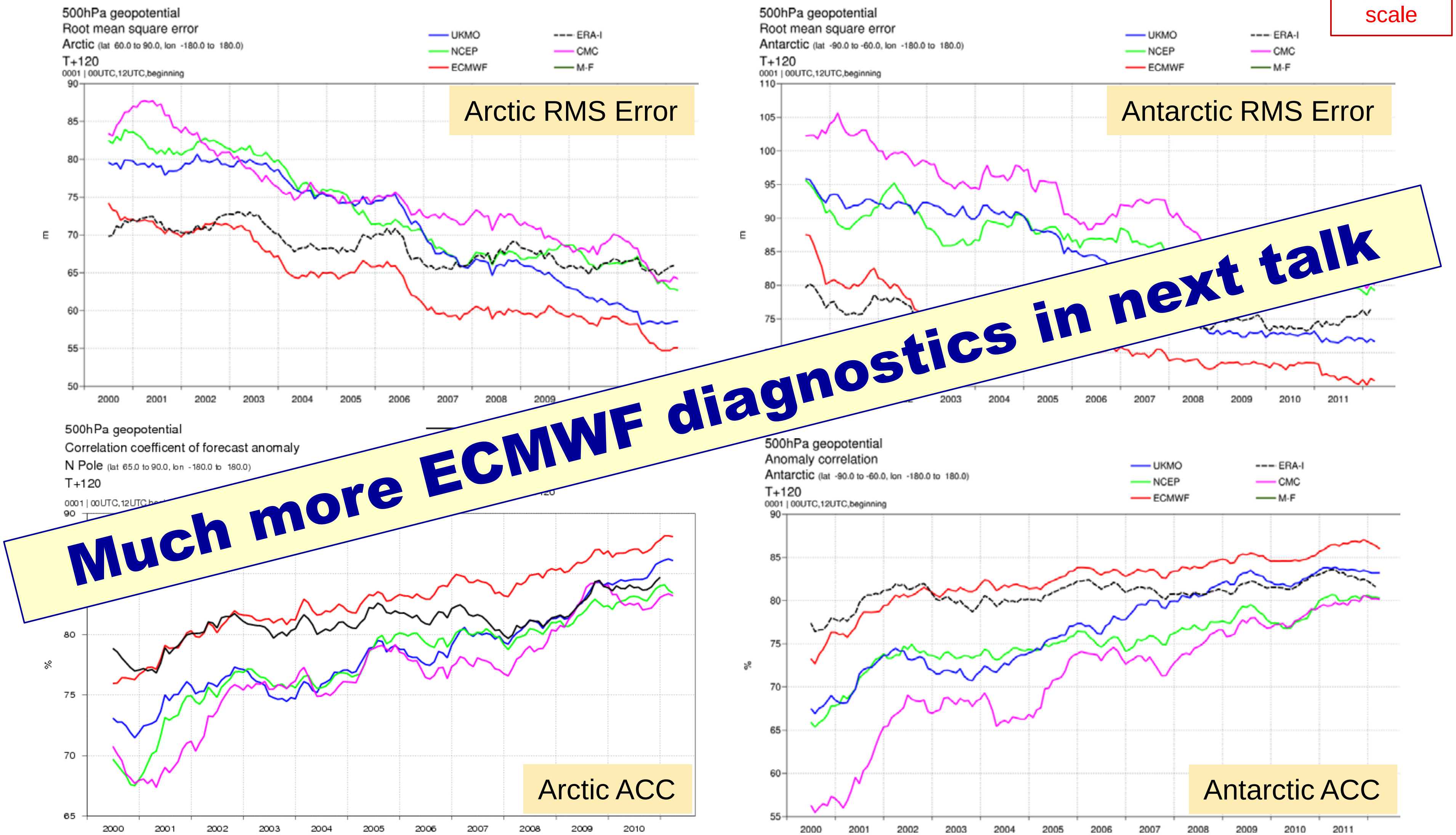
GA-4, GA-11

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

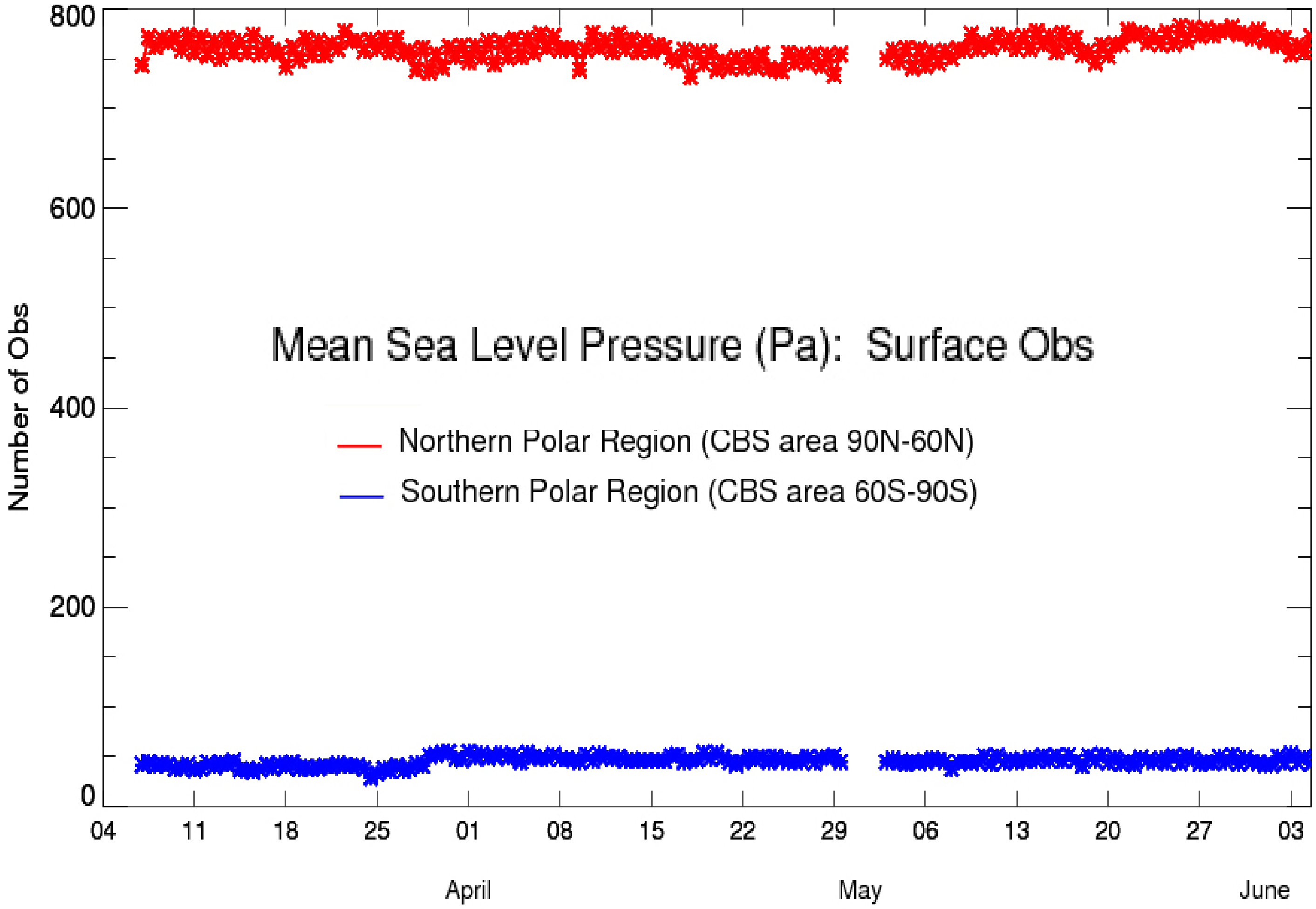


Comparison with other centres (2000-2012)

Note diff. y-axis scale



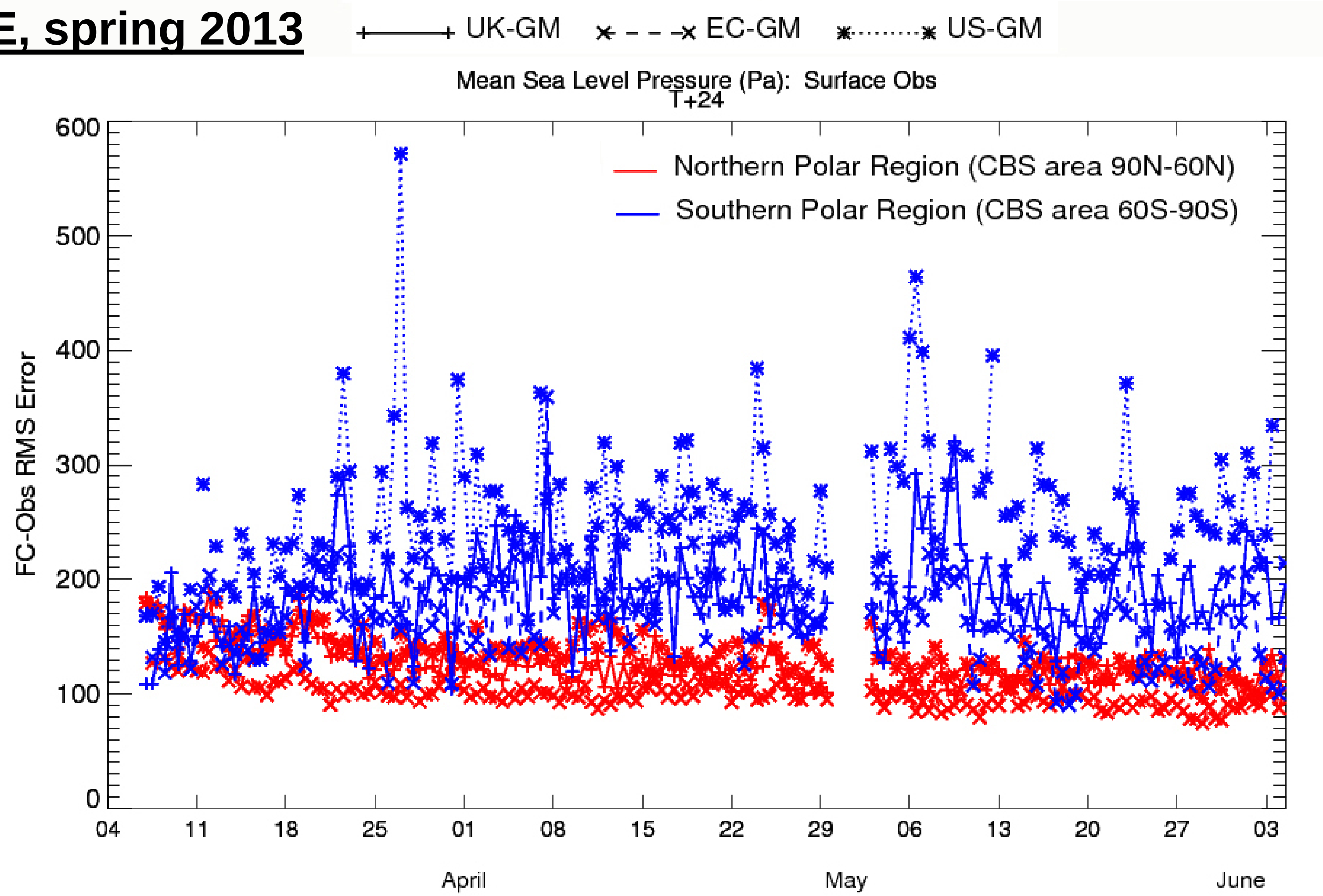
Obs availability, spring 2013



(Acknowledgement: Marion Mittermaier)

North - South Comparison

RMSE, spring 2013

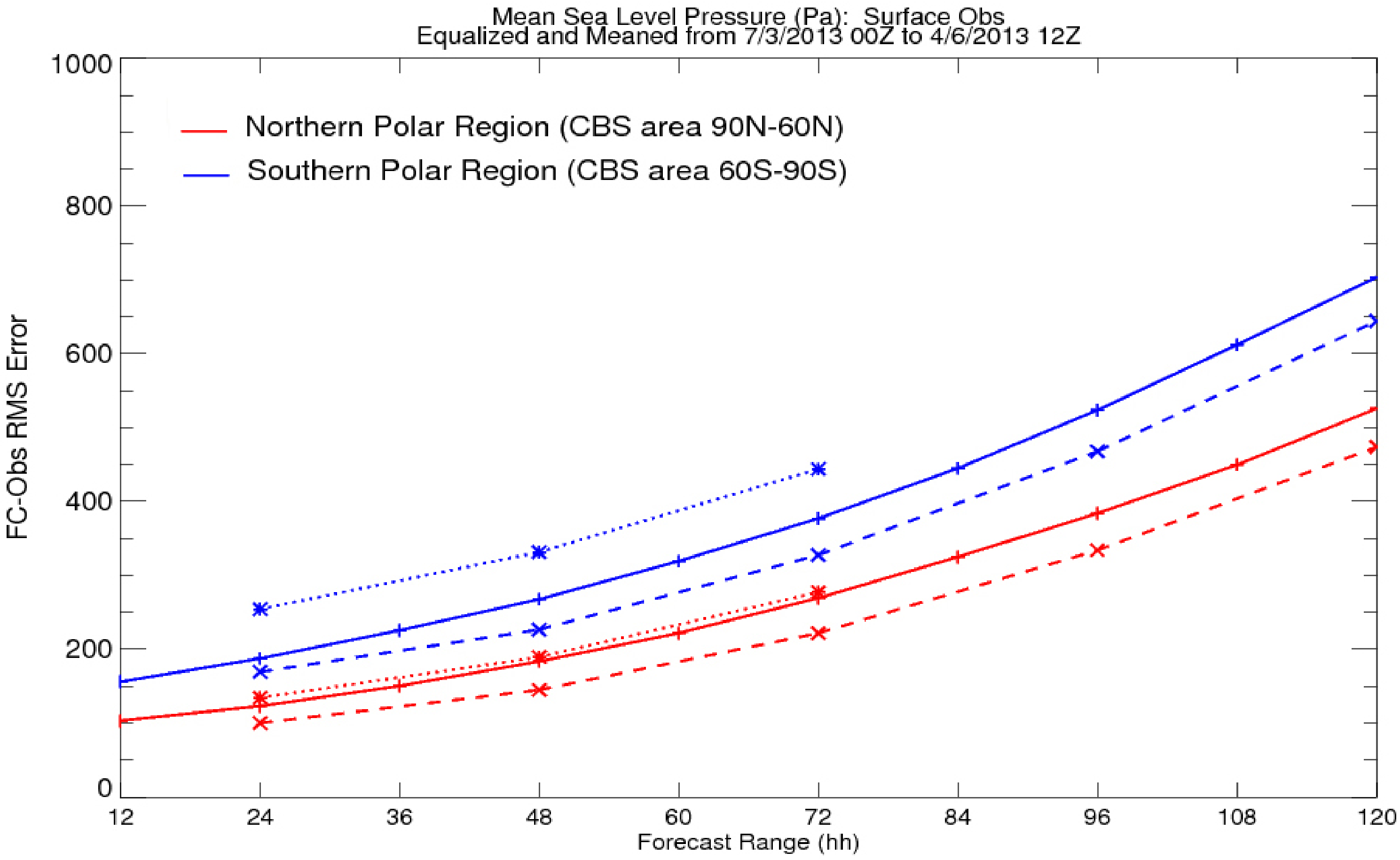


NB: UK @ full resolution; EC & US @ coarser CBS grip resolution

(Acknowledgement: Marion Mittermaier)

North - South Comparison

RMSE, spring 2013 UK-GM EC-GM US-GM **3-month mean**



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**

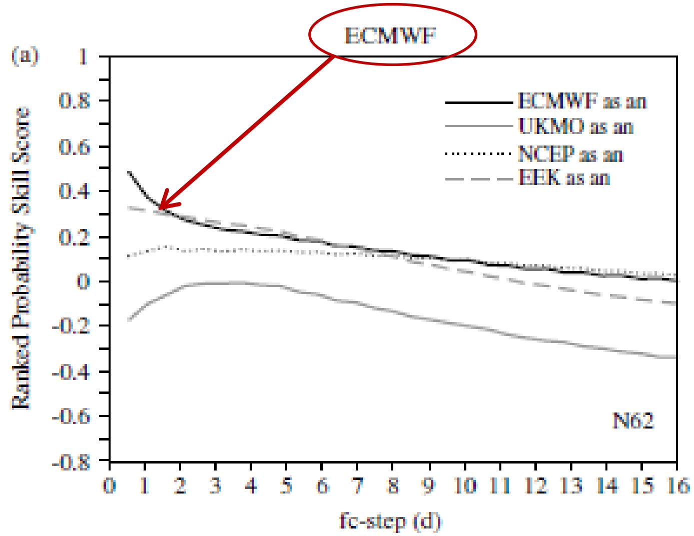
⬄ Utmost care if using model analysis as verifying *"truth"*

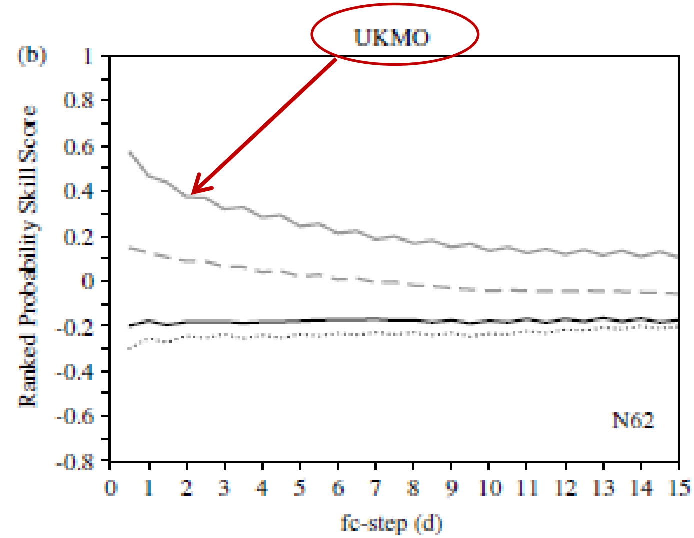
⬄ **Analyses typically are highly model dependent!**

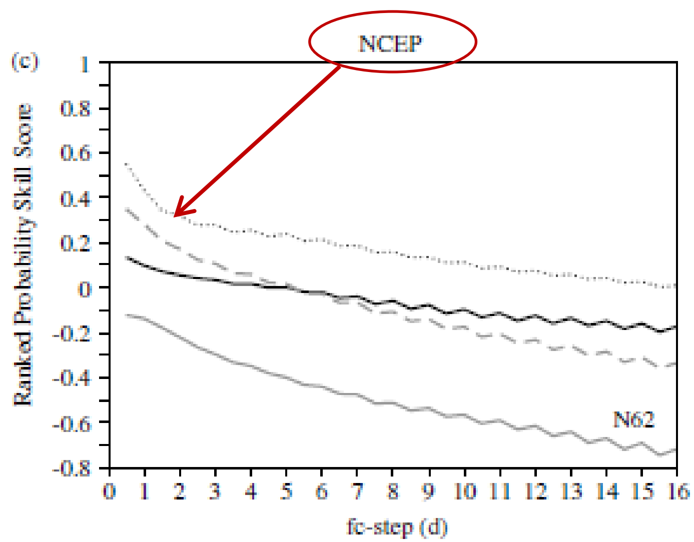
✓ **Especially so in polar regions with lack of observations**

⬄ 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 !







RPSS for 850 hPa temperature in the tropics (TIGGE data)
(from Park et al, 2008)

You always get best verification scores when using your <u>own</u> analysis

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

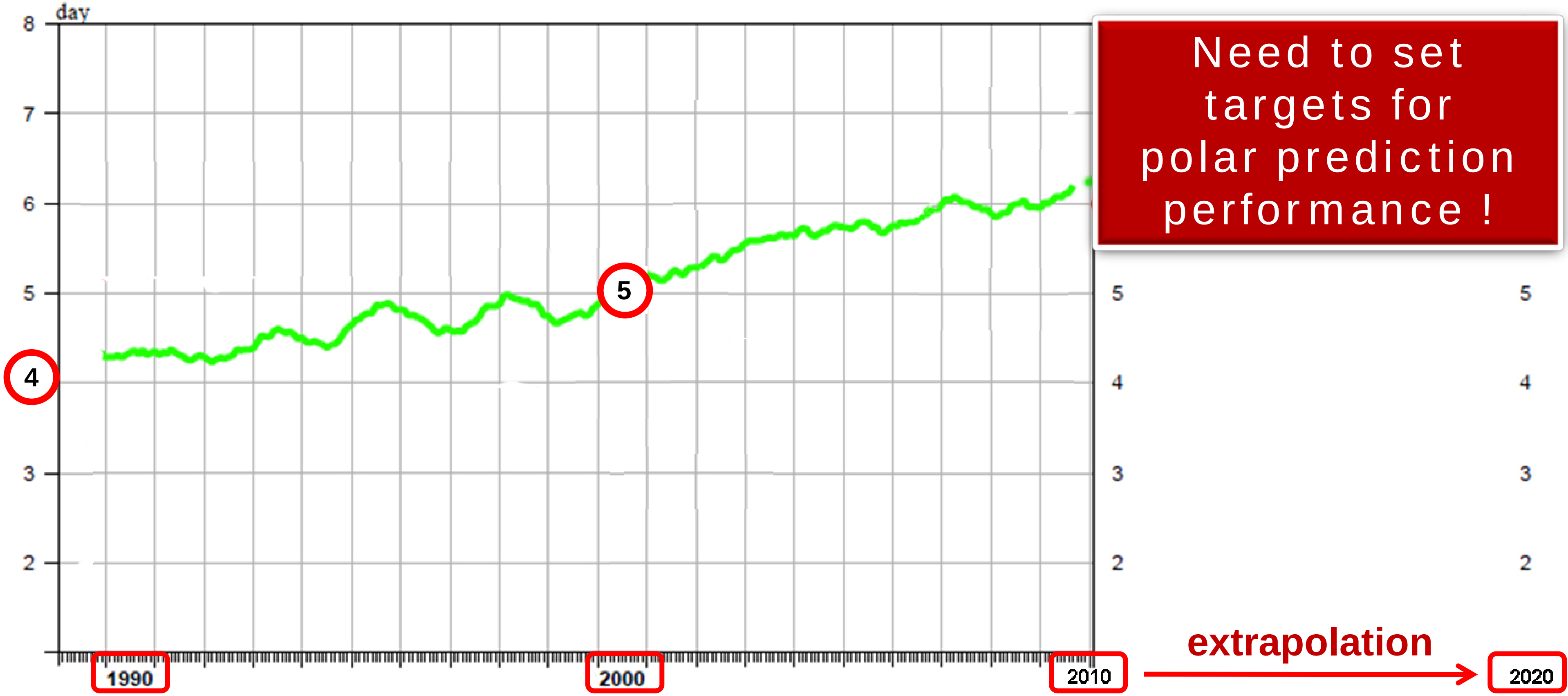
Repeat this kind of experiment for the Polar regions

Predictability - Free atmosphere

Predictability ⇔ ECMWF "headline" measure

ACC of Z 500 remains above 80 %

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



Need to set targets for polar prediction performance !

pertti.nurmi@fmi.fi

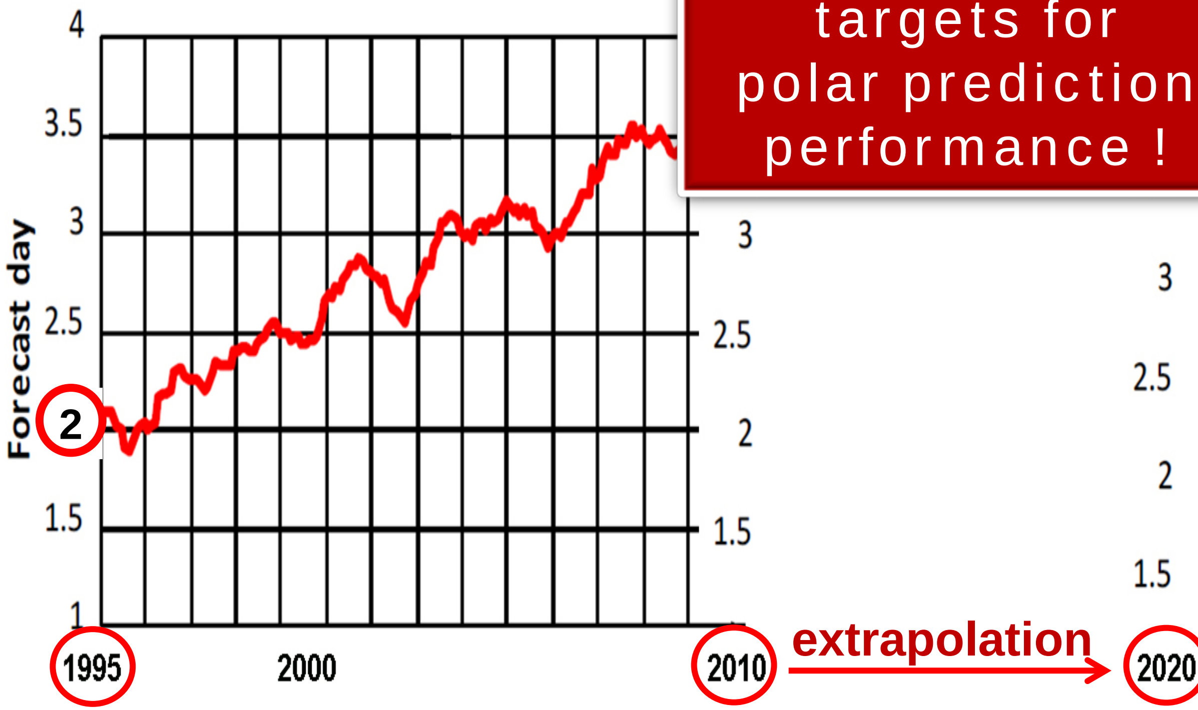
Predictability - Surf. weather ⇔ End-user perspective

Predictability ⇔ ECMWF "headline" measure

"1 – SEEPS" of 24 hr precipitation remains above 45 %

- ✓ **1995 ⇔ 2 days**
- ✓ **2010 ⇔ 3.5 days** ⇔ Expected increase ⇔ 1 day / decade

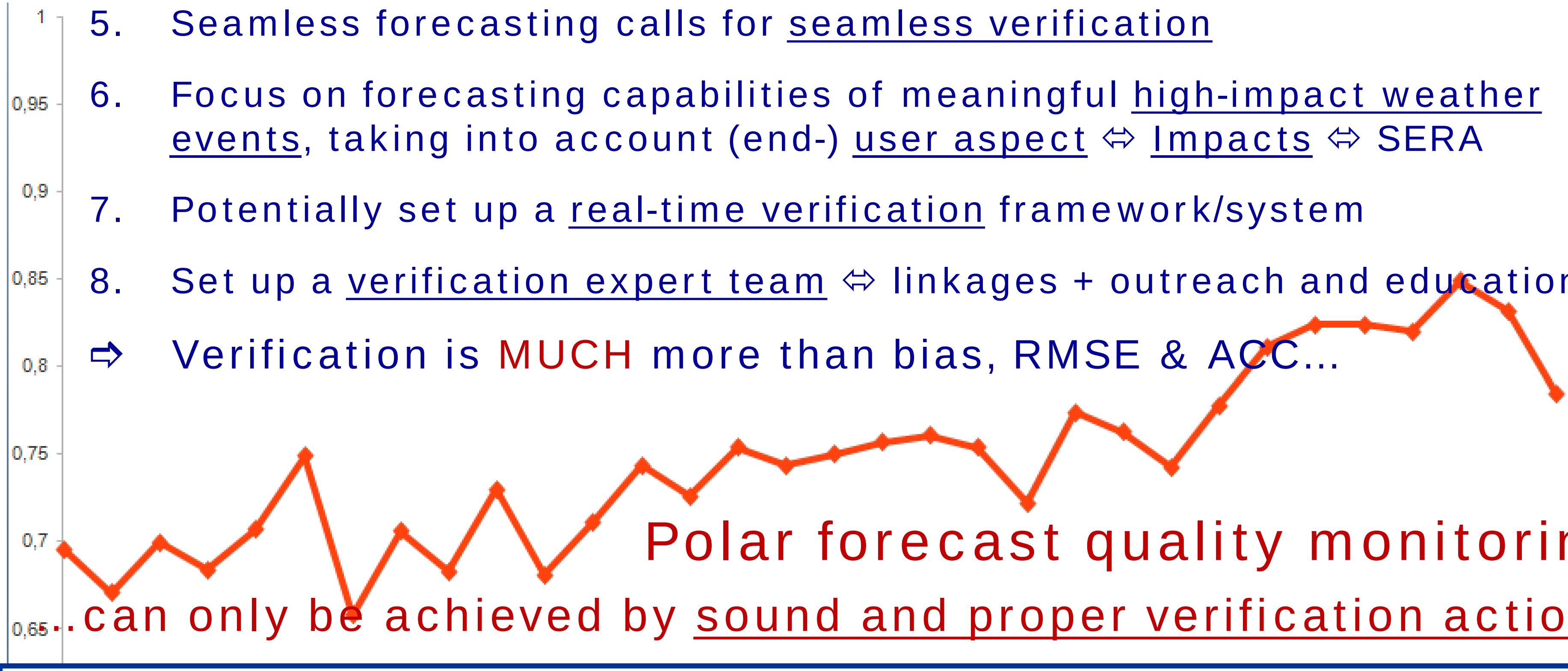
Need to set targets for polar prediction performance !



“ Summary “

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...



Polar forecast quality monitoring

...can only be achieved by sound and proper verification actions

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


Preparation Phase
2012-2016
YOPP
2017-2018
Consolidation
Phase
2018-2022
JWGFVR
thanks
You
Pertti
Laurie
Beth
© Timo Vihma