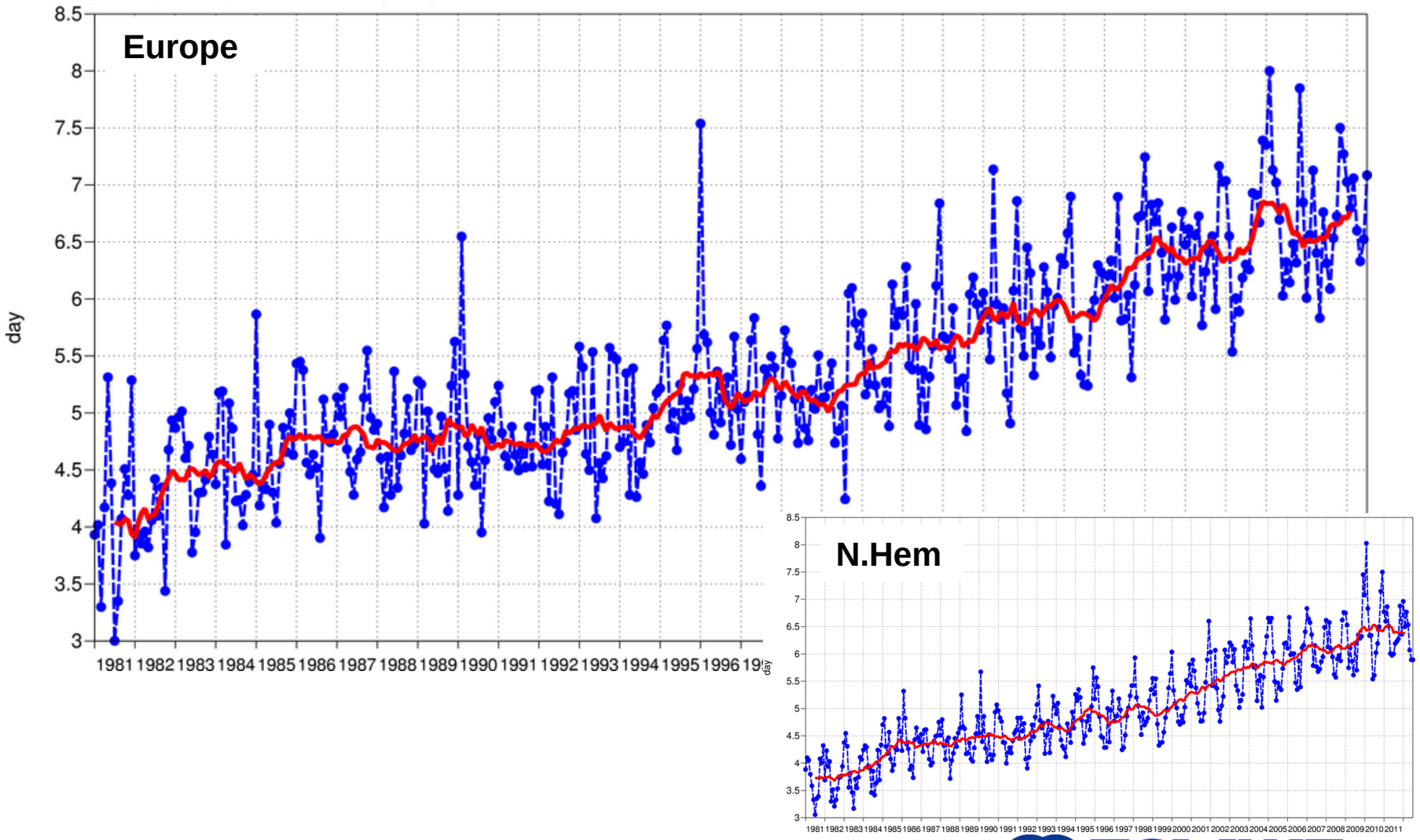
Verification scores including polar verification

Jean-Noël Thépaut - ECMWF

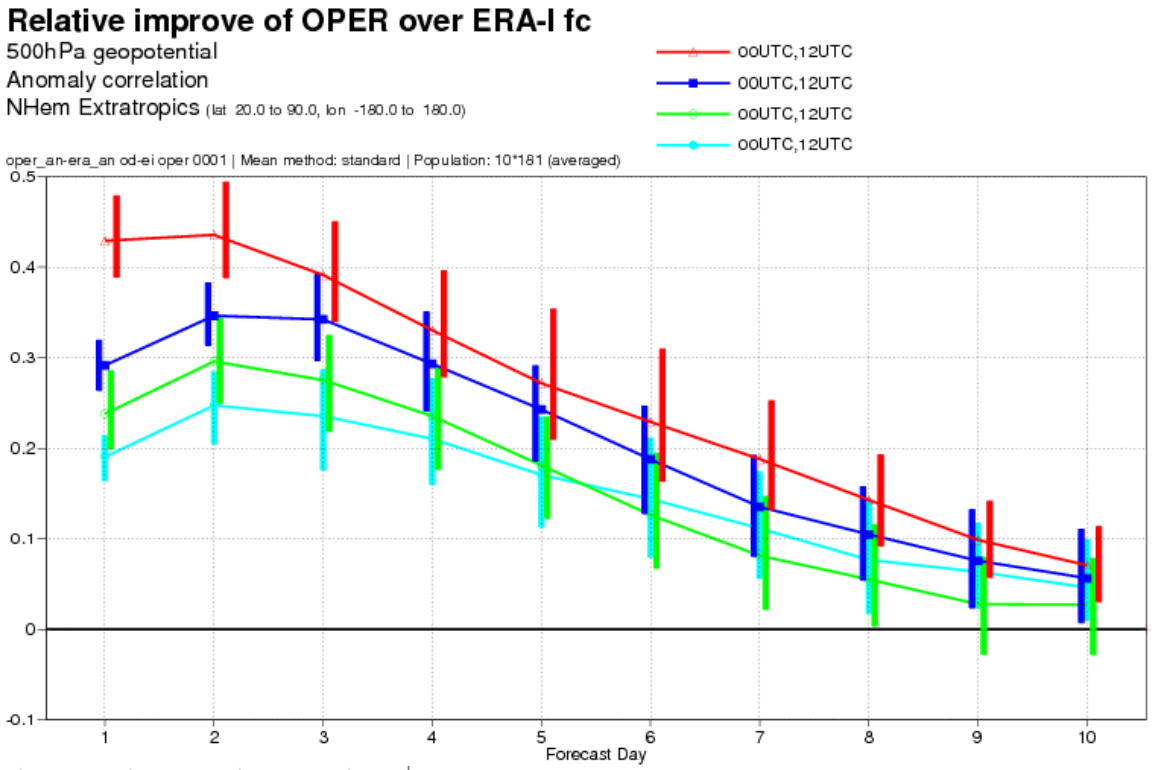
- Scores evolution between 2011 and 2012
- ECMWF: WMO Lead Centre for Deterministic Forecast Verification
- Polar verification
  - Versus analysis and versus obs
  - Some statistics based on Concordiasi dataset

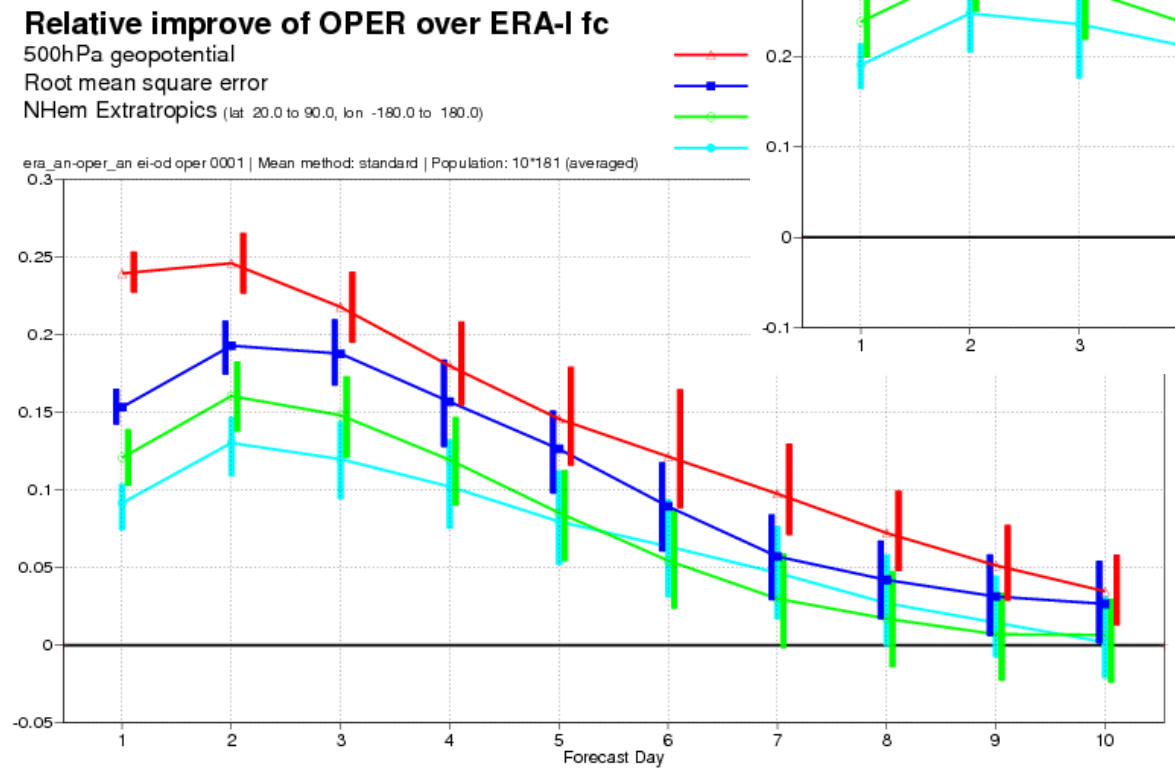
Acknowledgements: Matin Janousek, David Richardson, Vincent Guidard

Primary Headline Score
Z500, Time series of ACC=0.8 Europe

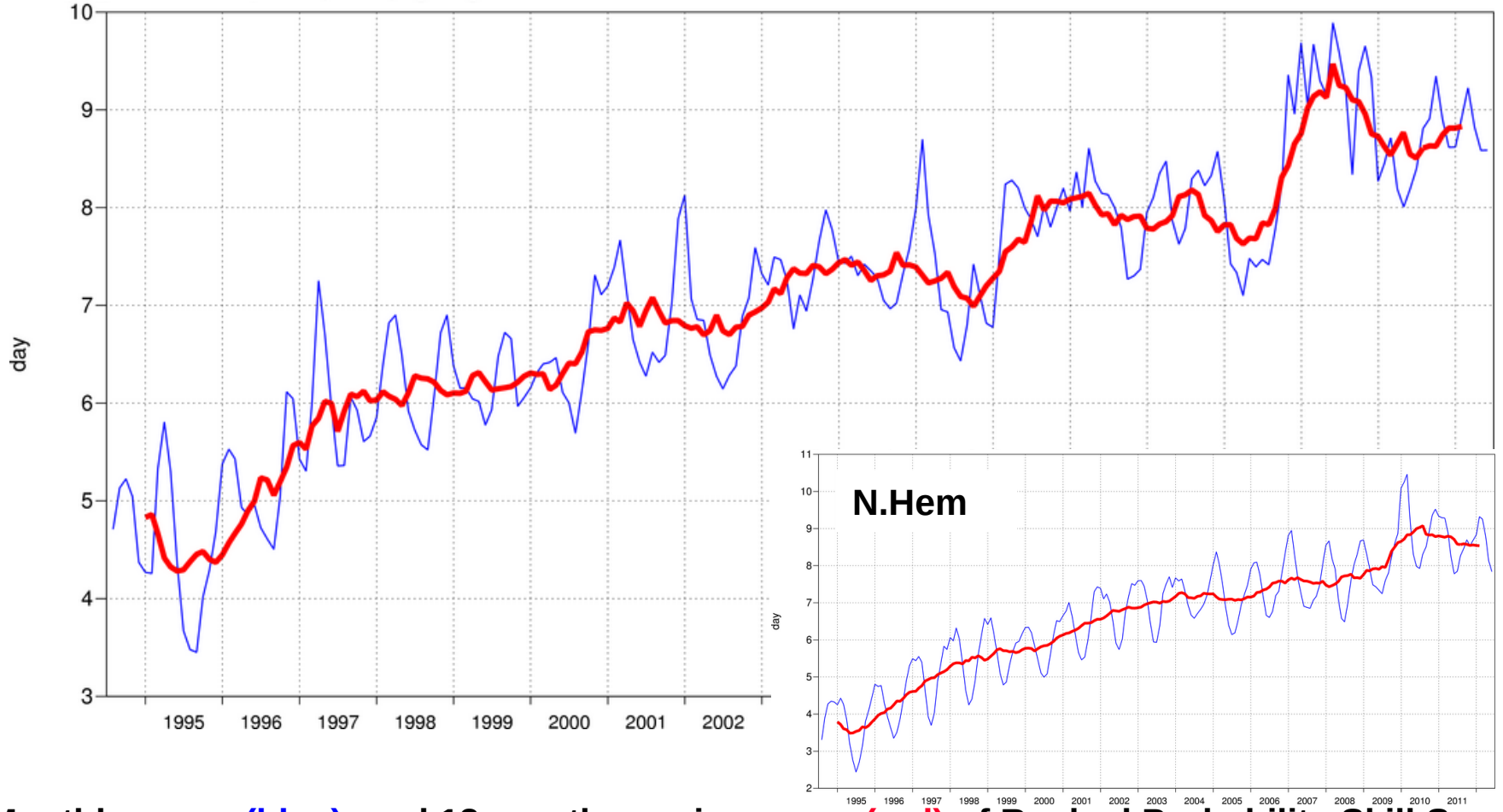


High-res v ERA-I N hem





Primary Headline Probabilistic Score
RPSS, T850 Europe



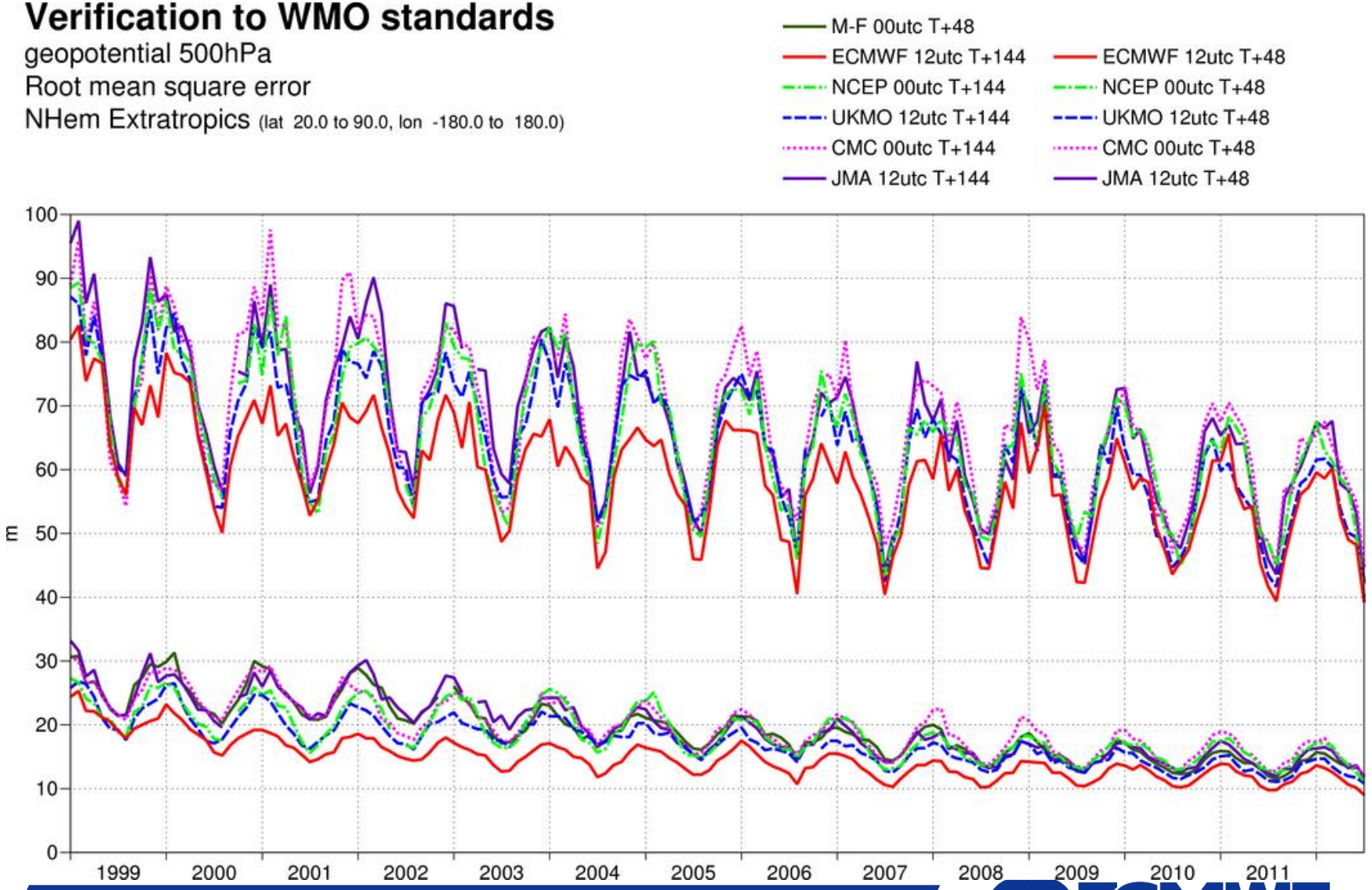
Monthly score (blue), and 12-month running mean (red) of Ranked Probability Skill Score for EPS forecasts of T850 hPa for Europe. Day at which score reaches 25%.

ECMWF

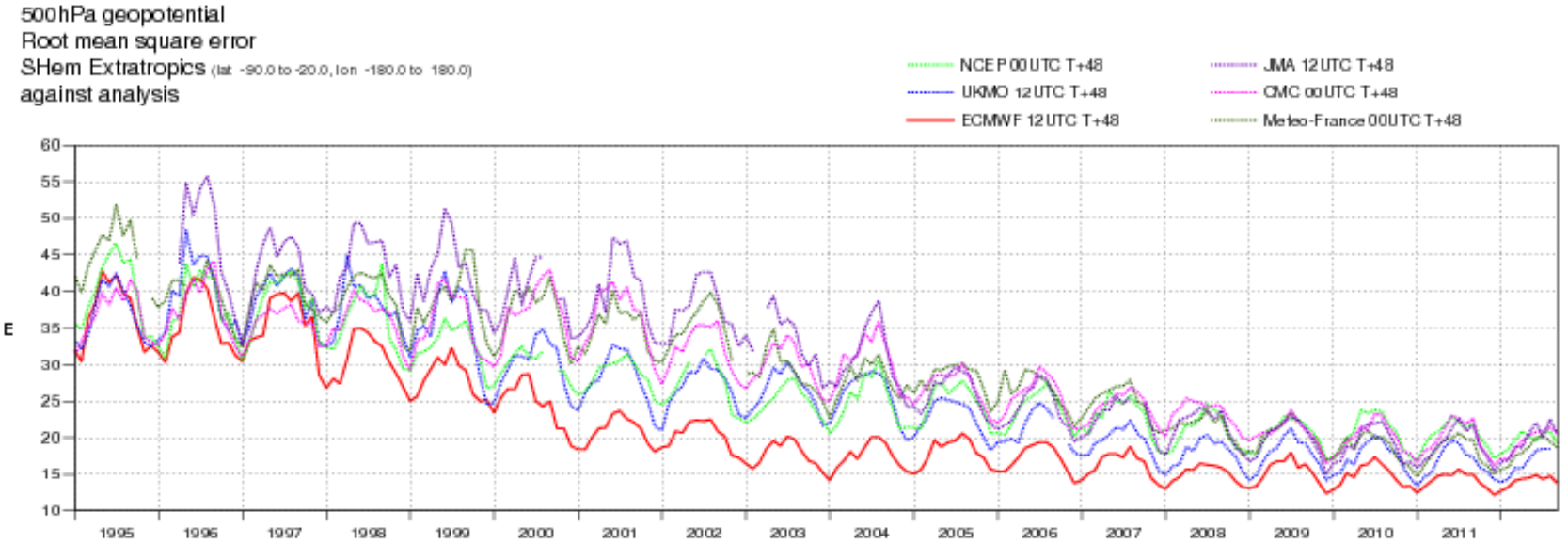
WMO scores Z500 N.Hem

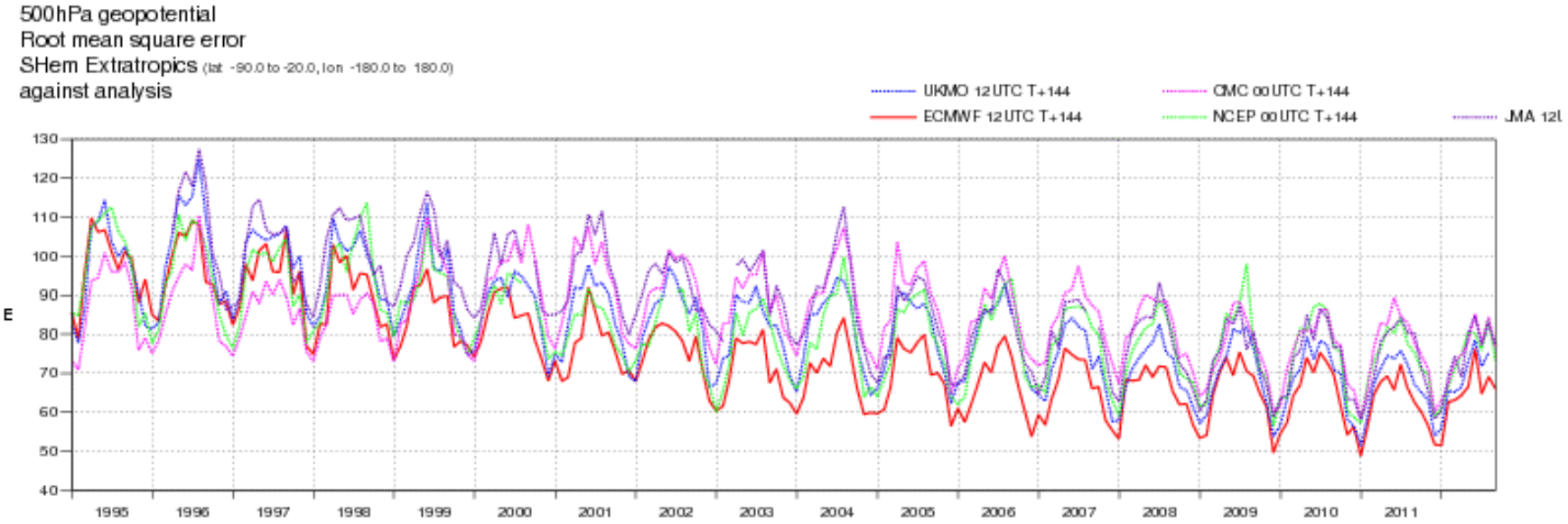
Verification to WMO standards

geopotential 500hPa
Root mean square error
NHem Extratropics (lat 20.0 to 90.0, lon -180.0 to 180.0)

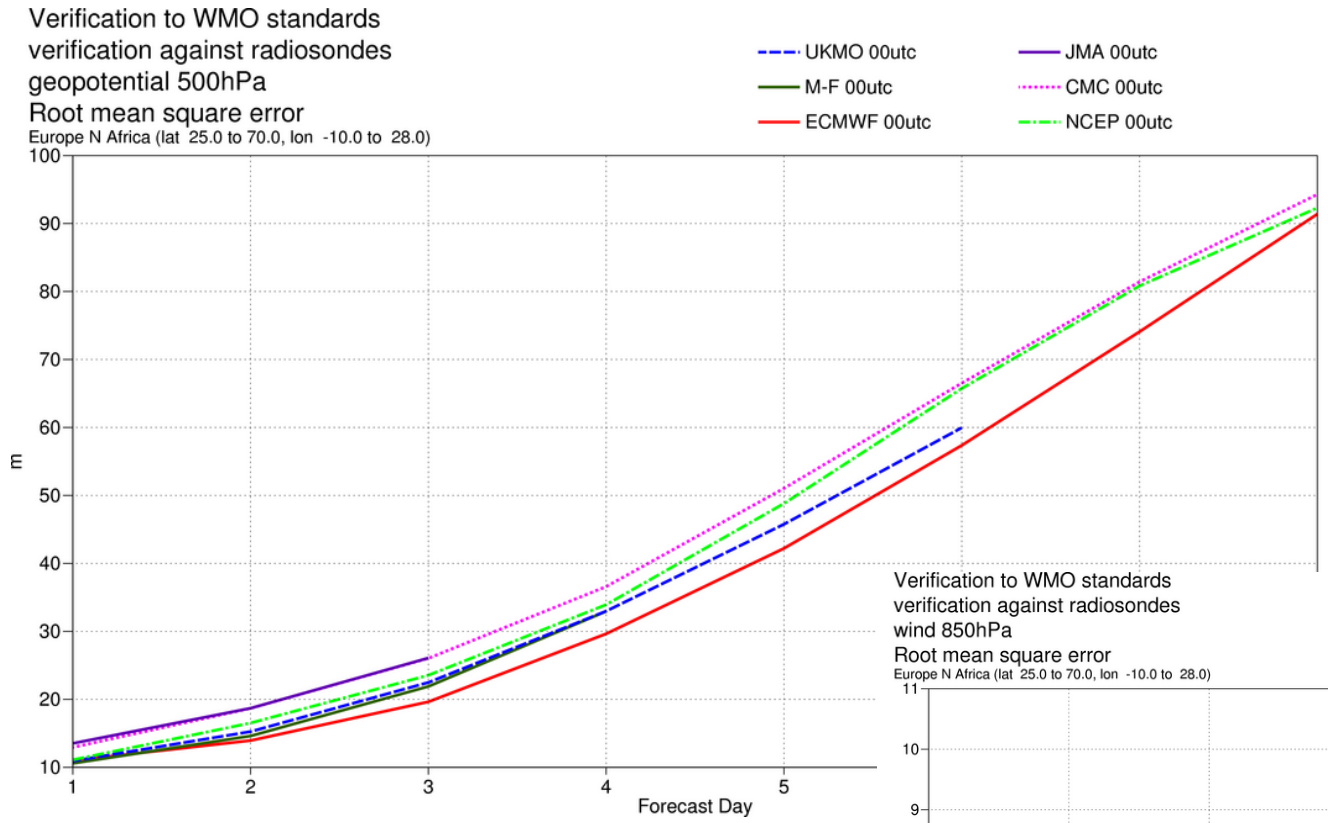


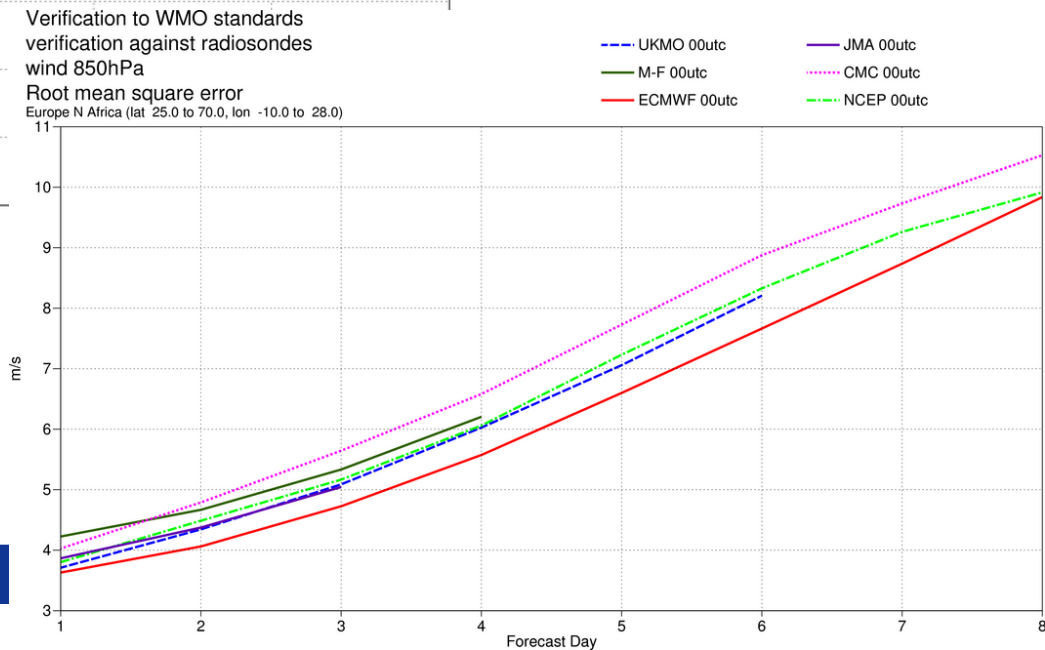
WMO scores Z500 S.Hem





WMO scores using radiosondes Z500 and wind850 over Europe





2 day fcst error against observations - NH

6 day fcst error against observations - NH

2 day fcst error against observations - SH

6 day fcst error against observations - SH

NCEP stands out of other centres

and now, the demo...

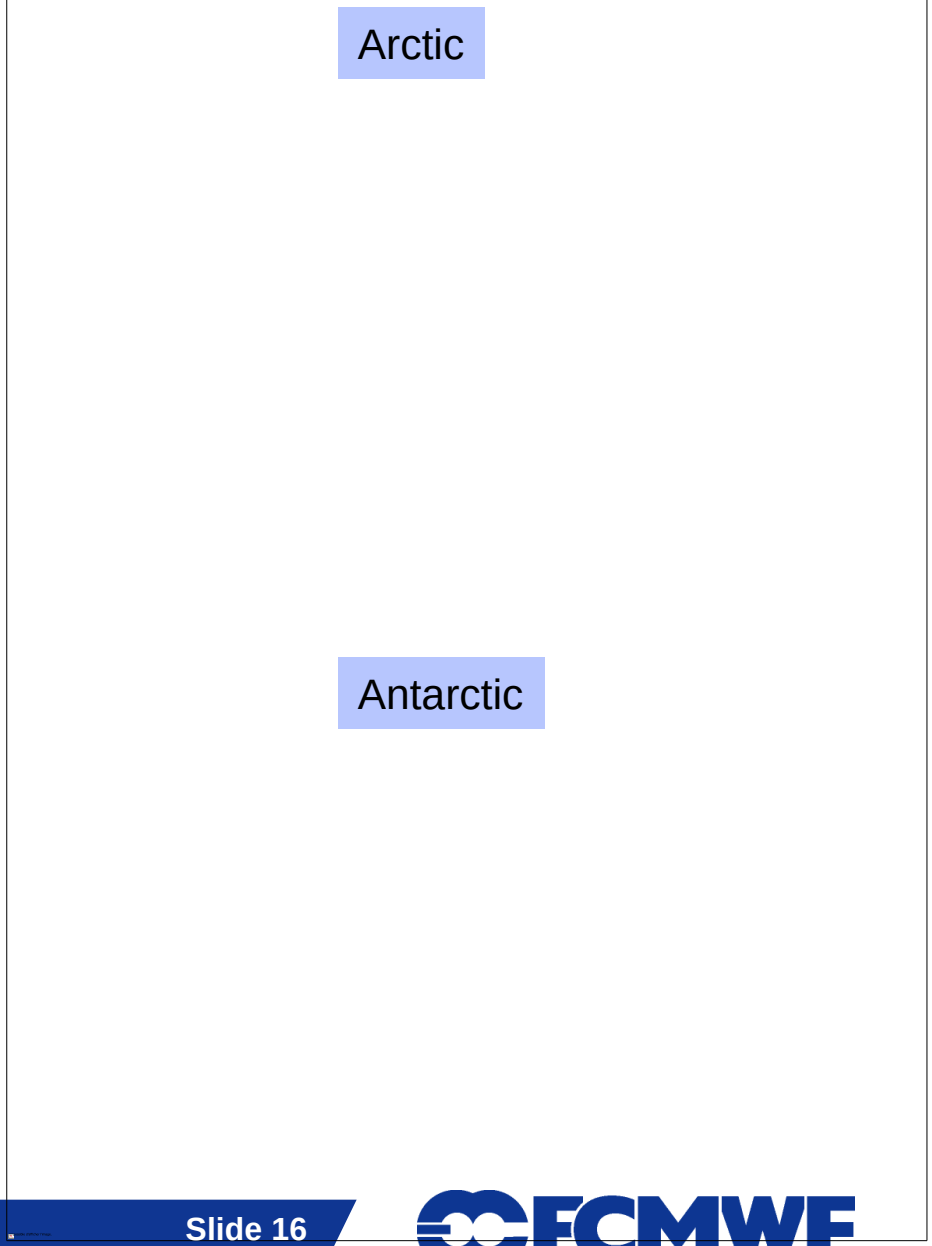
http://apps.ecmwf.int/wmolcdnv/

Verification for polar regions (M. Janousek, D. Richardson)

- Scores computed for polewards of 60°
- Verification at ECMWF using available fields from other centres
- Done for Z500 and T850
- All verification against analysis (each centre against own analysis) or radiosonde observations
- ERA-Interim scores shown as reference (ERA is fixed model and assimilation system)

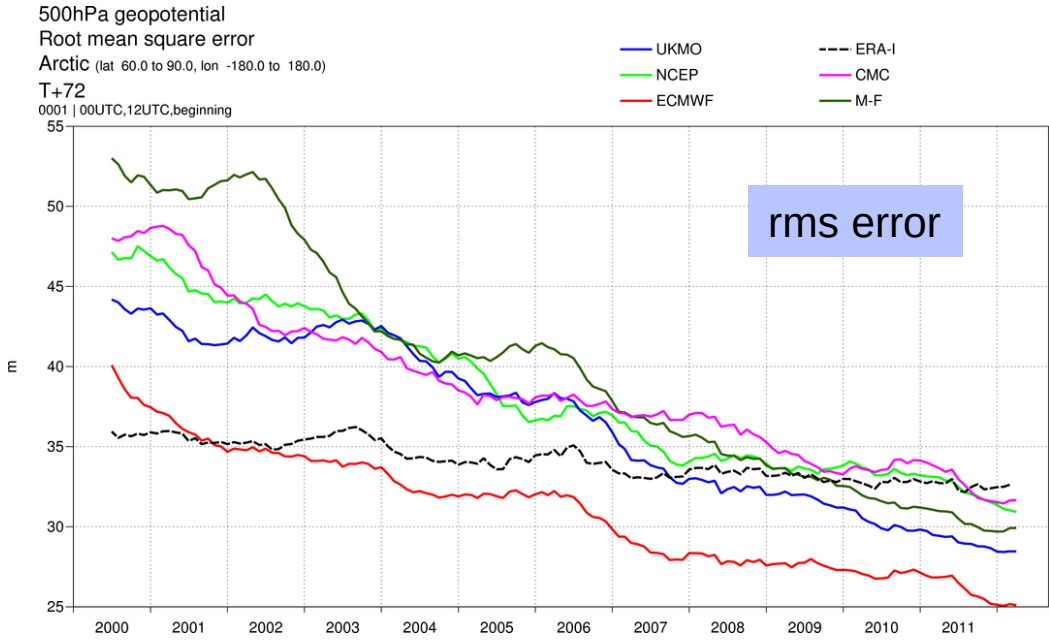
ECMWF operational and ERA-Interim (1990-2012)

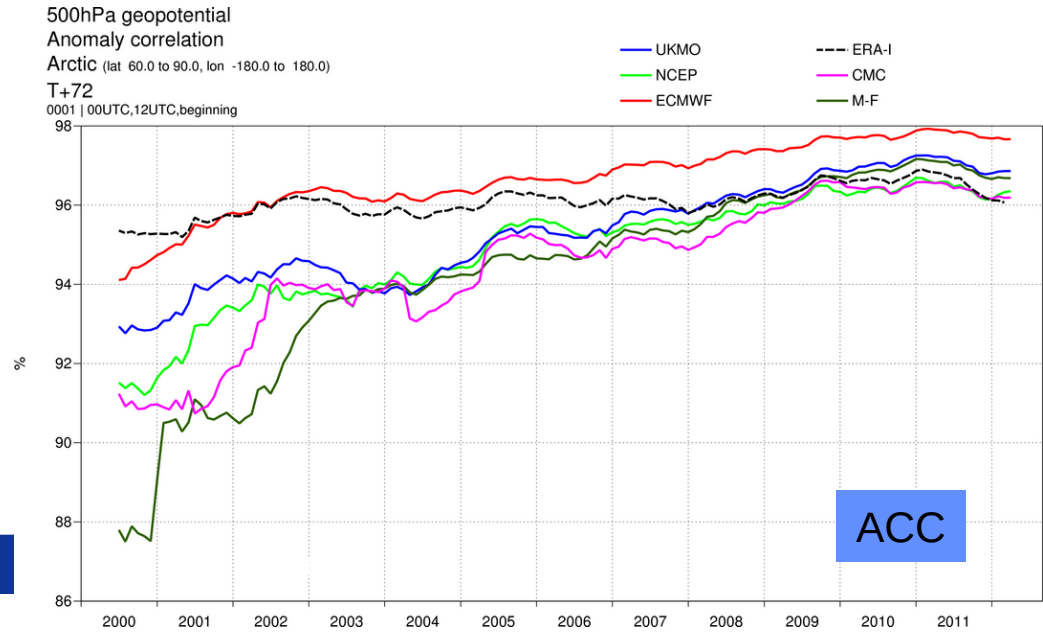
- **Z500 ACC=80%, 12-month moving average**
- **Arctic: clear improvement in system around 2000, and consistently better than ERA beyond 2002. But the apparent change 2001-2002 and 2008-09 are matched in ERA. Drop in skill and predictability in 2012.**
- **Antarctic: clear sustained improvement in 1990s; still positive trend**
- **ERA changes: either atmospheric variability or changes to observing system**



Comparison with other centres (2000-2012) Arctic

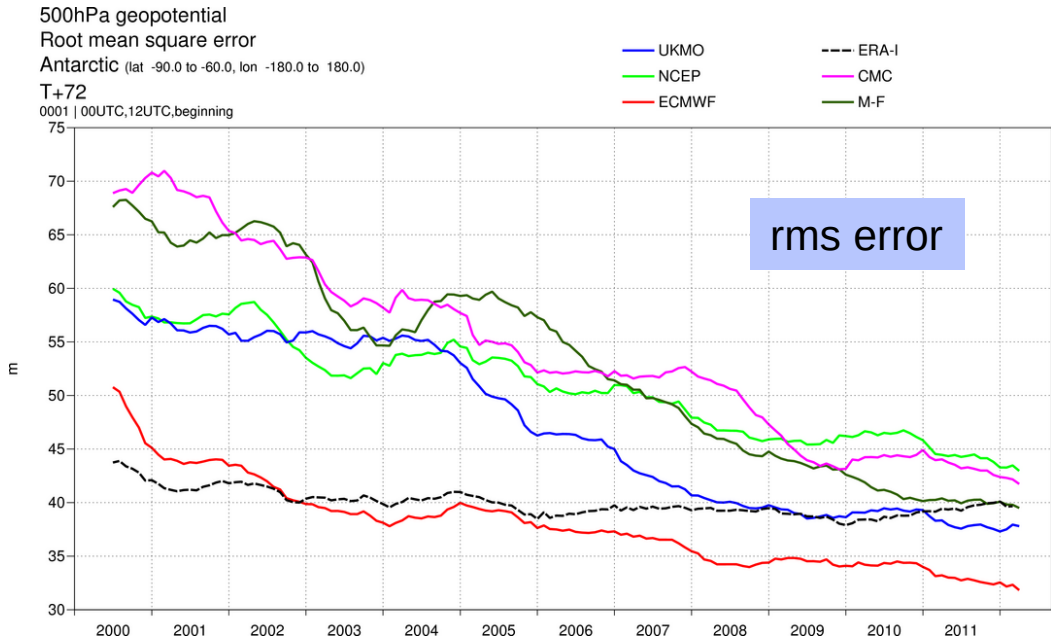
- Day 3 forecasts (T+72)
- Z500, 12-month moving average
- Each centre verified against own analysis
- ERA-I shown for reference

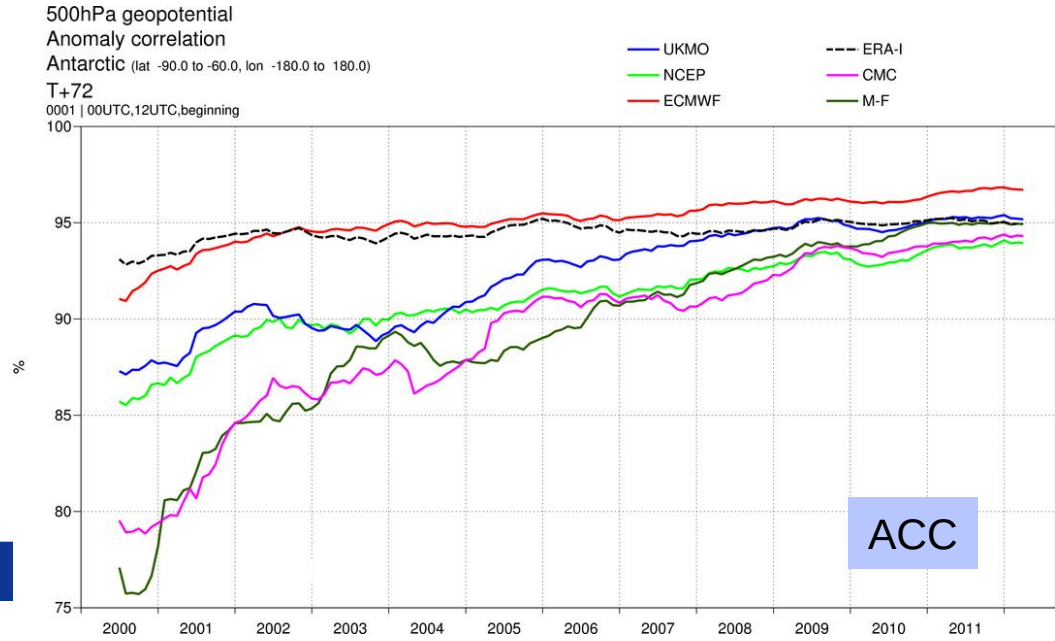




Comparison with other centres (2000-2012) Antarctic

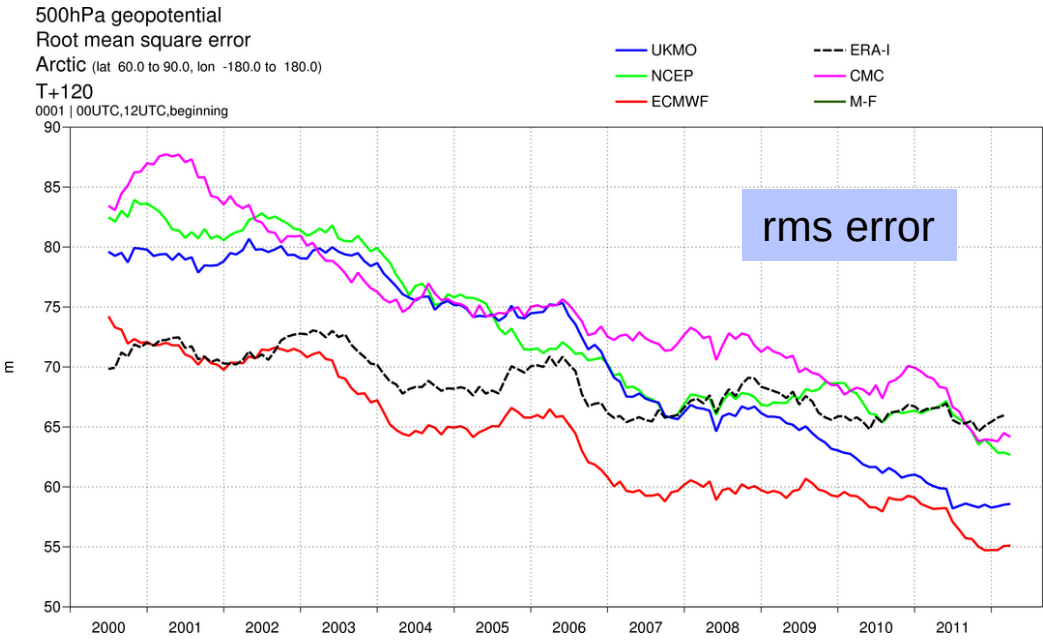
- Day 3 forecasts (T+72)
- Z500, 12-month moving average
- Each centre verified against own analysis
- ERA-I shown for reference

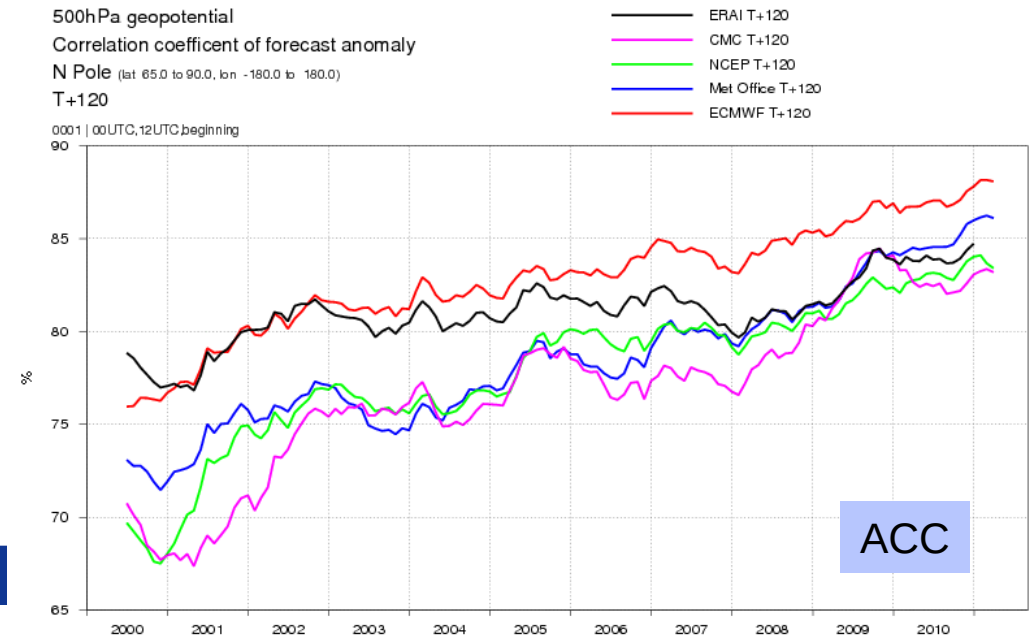




Comparison with other centres (2000-2012) Arctic

- Day 5 forecasts (T+120)
- Z500, 12-month moving average
- Each centre verified against own analysis
- ERA-I shown for reference

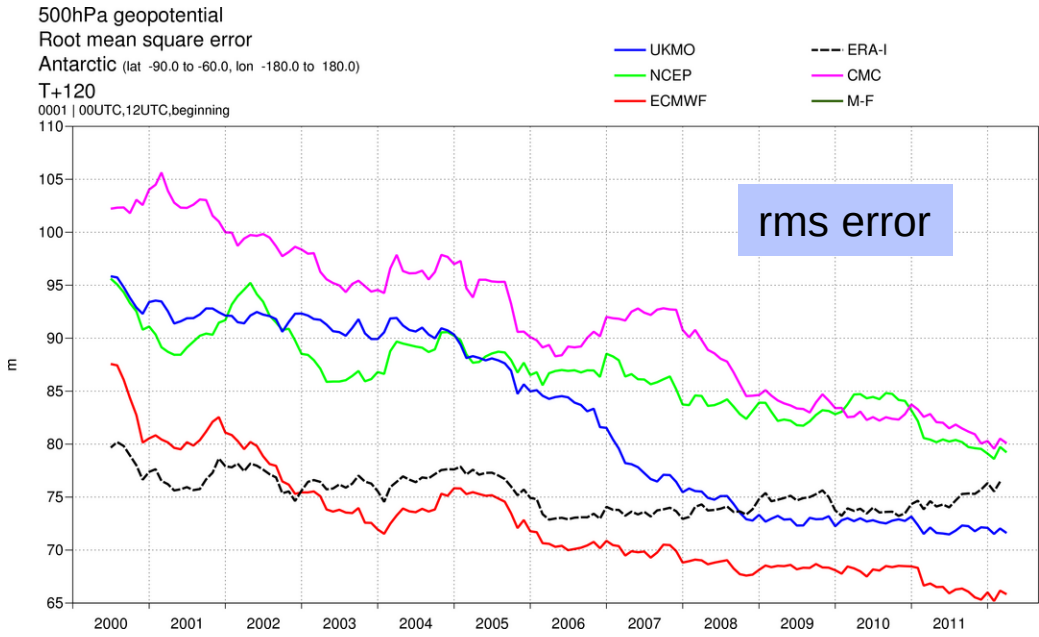


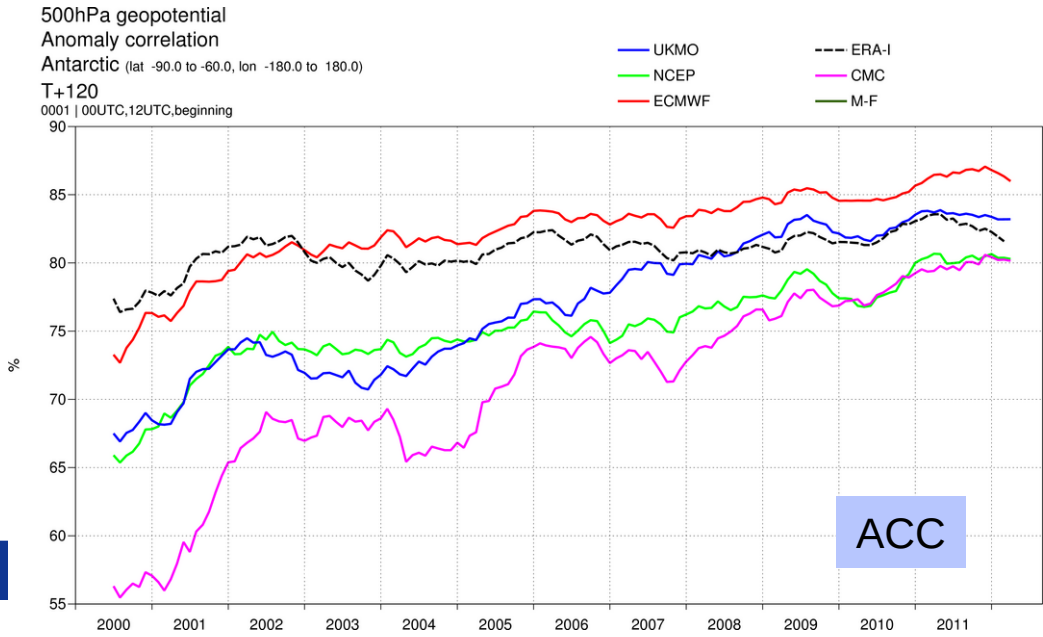


Comparison with other centres (2000-2012) Antarctic

- Day 5 forecasts (T+120)
- Z500, 12-month moving average
- Each centre verified against own analysis
- ERA-I shown for reference

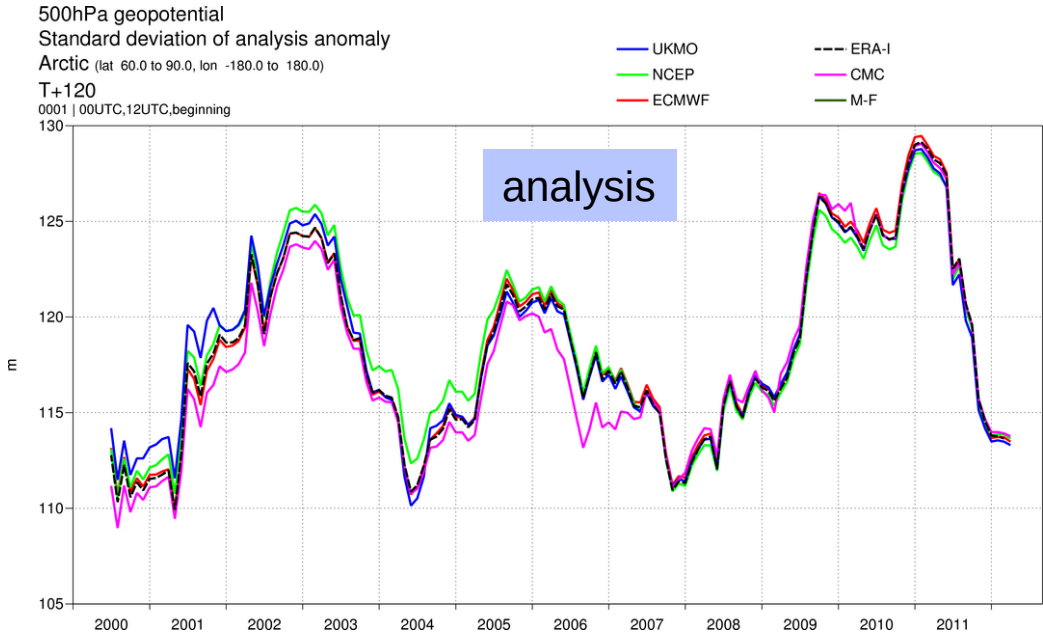
- NB some dates missing for CMC in 2009 – affects these scores for 2009 (other years OK)

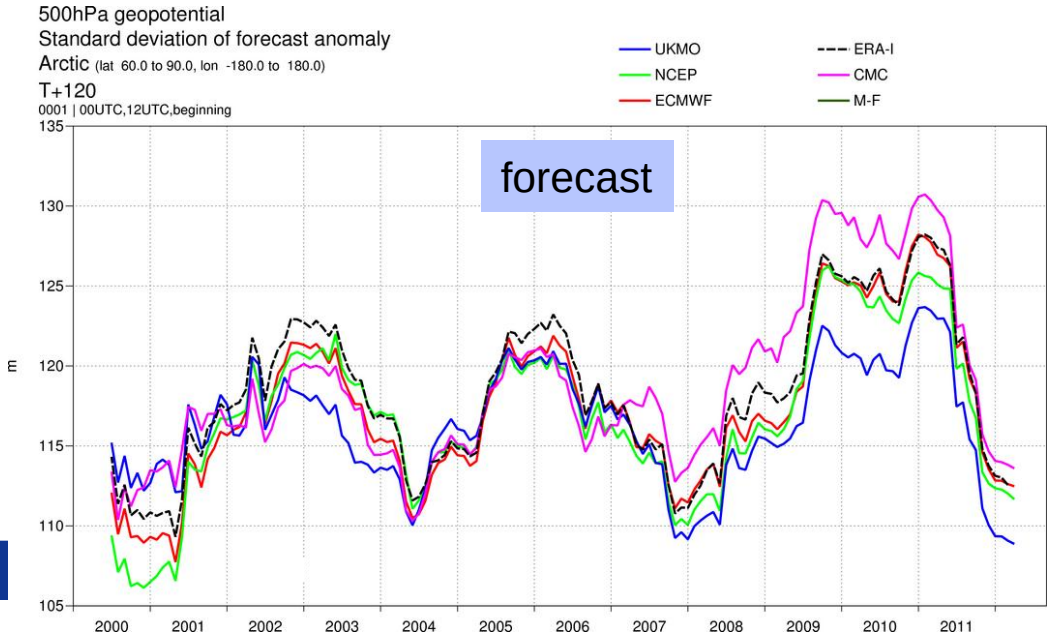




Comparison with other centres (2000-2012) Arctic

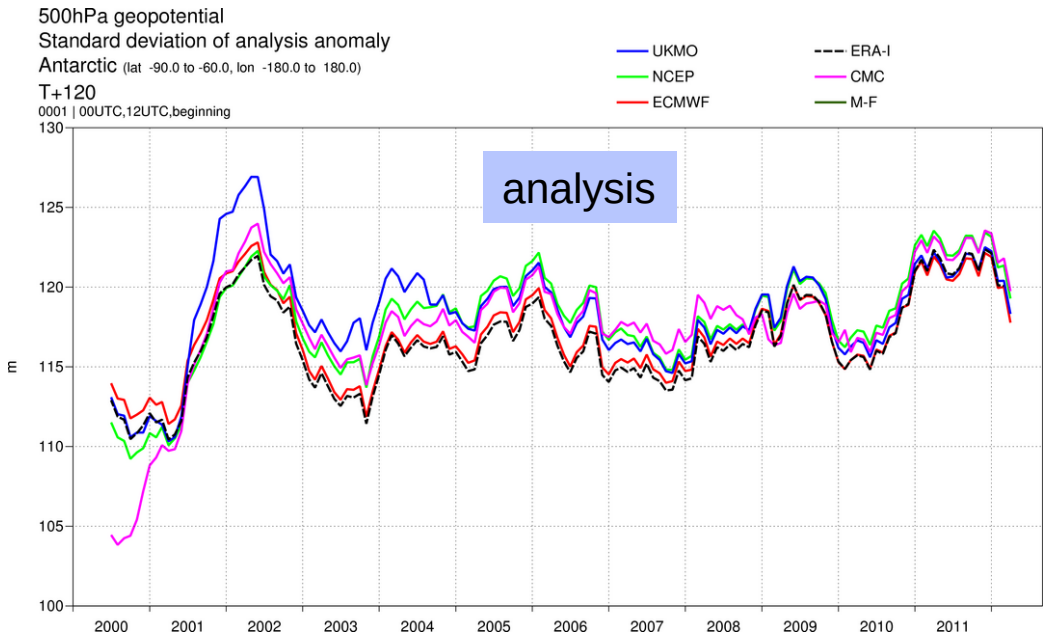
- Variability (activity) of forecast and analysis fields: standard deviation of anomalies
- Day 5 forecasts (T+120)
- Z500, 12-month moving average
- ERA-I shown for reference
- Compared to the analysis, Met Office forecast now rather underactive; CMC overactive (this can affect the rms errors)
- Drop of activity in 2012
- NB some dates missing for CMC in 2009 – affects these scores for 2009 (other years OK)

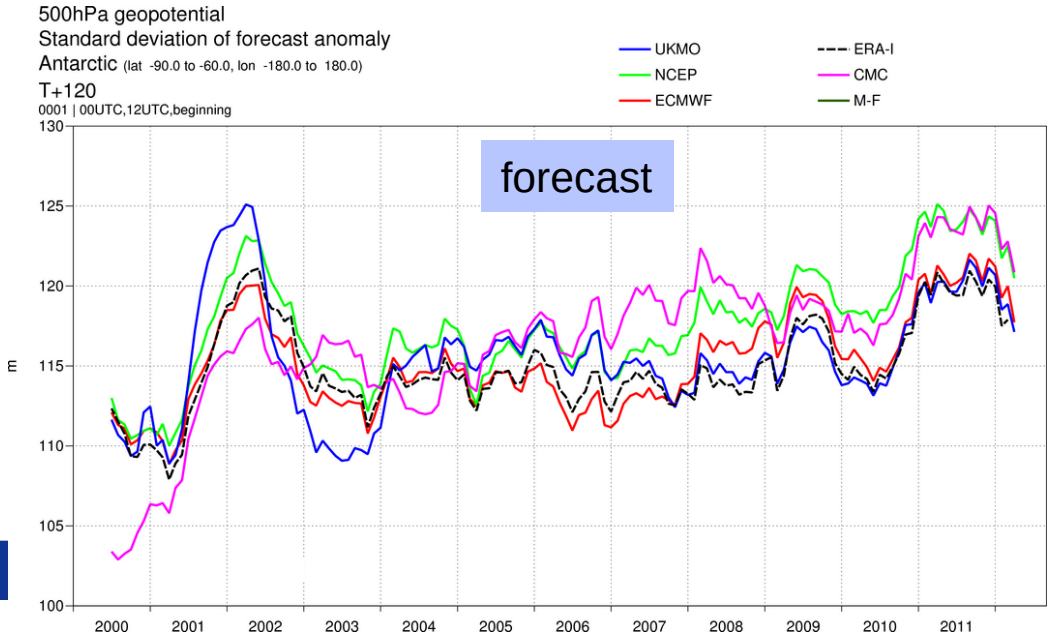




Comparison with other centres (2000-2012) Antarctic

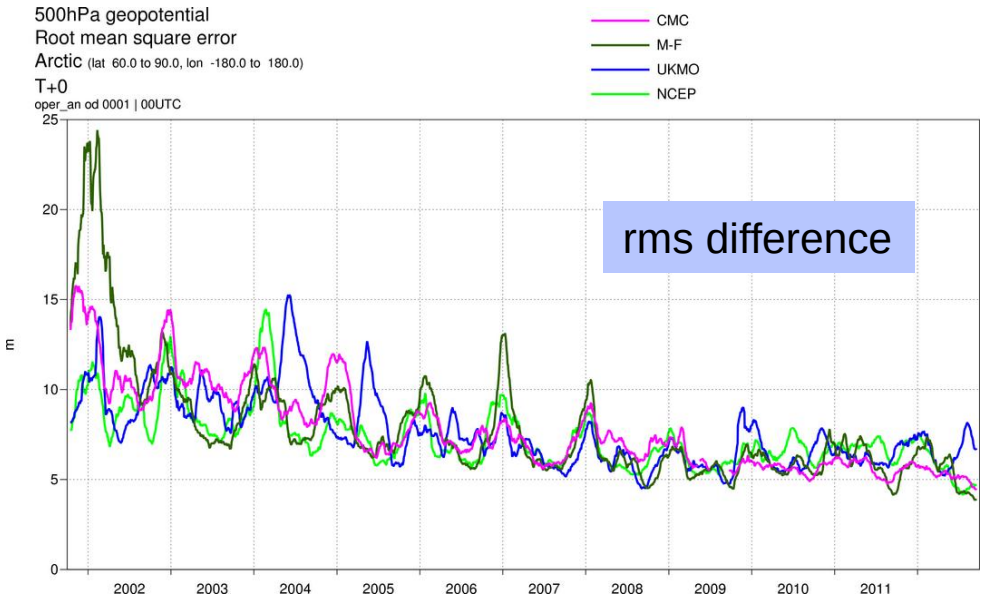
- Variability (activity) of forecast and analysis fields: standard deviation of anomalies
- Day 5 forecasts (T+120)
- Z500, 12-month moving average
- CMC and NCEP analyses more active than MetOffice and ECMWF
- CMC and NCEP overactive; MetOffice and ECMWF underactive (this can affect the rms errors)
- NB some dates missing for CMC in 2009 – affects these scores for 2009 (other years OK)

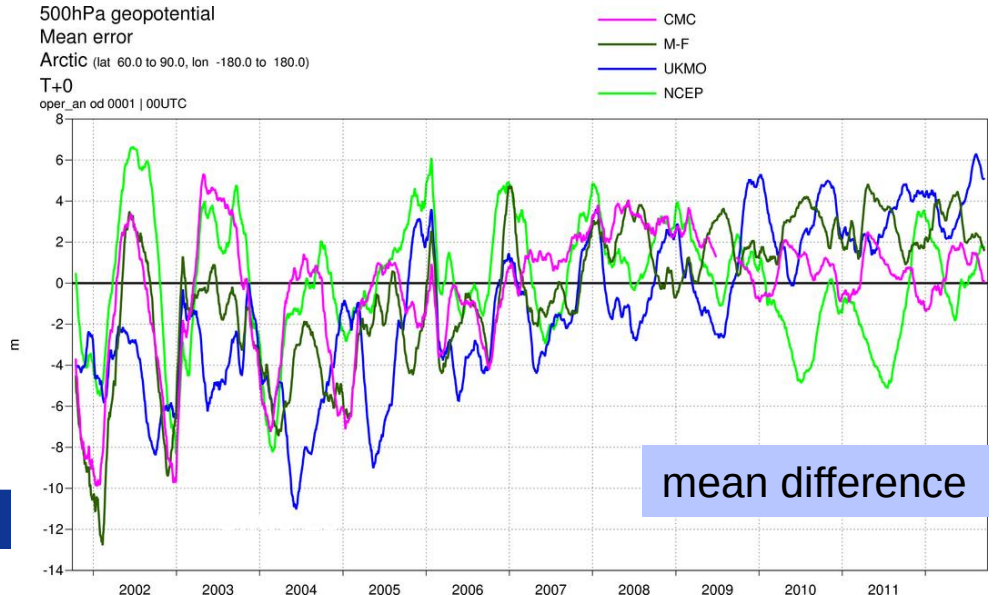




Comparison between analyses (Arctic)

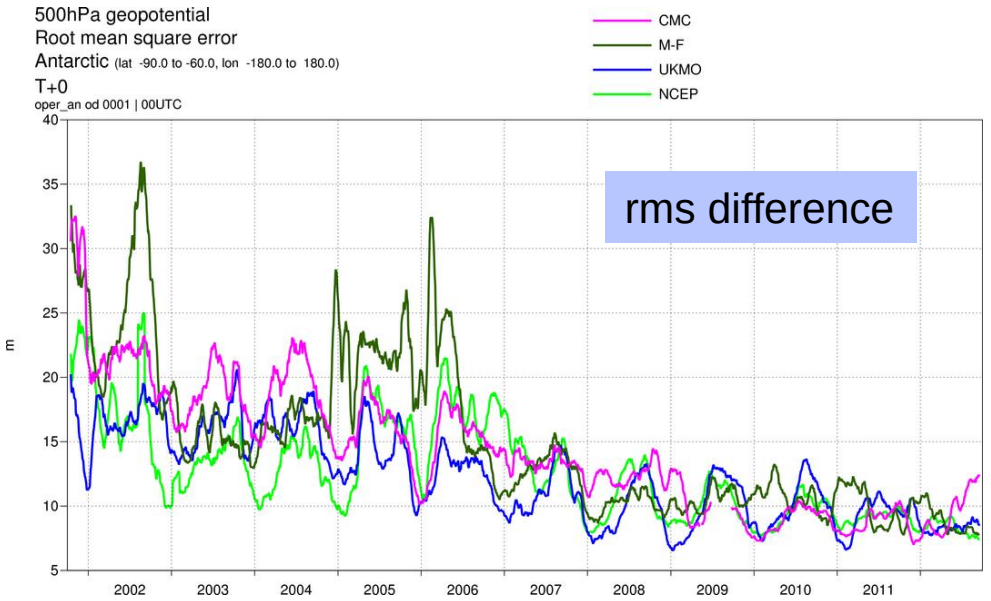
- Differences between the analyses of different centres
- Z500 30 day moving average
- Decrease over last decade in the difference between the analyses of different centres

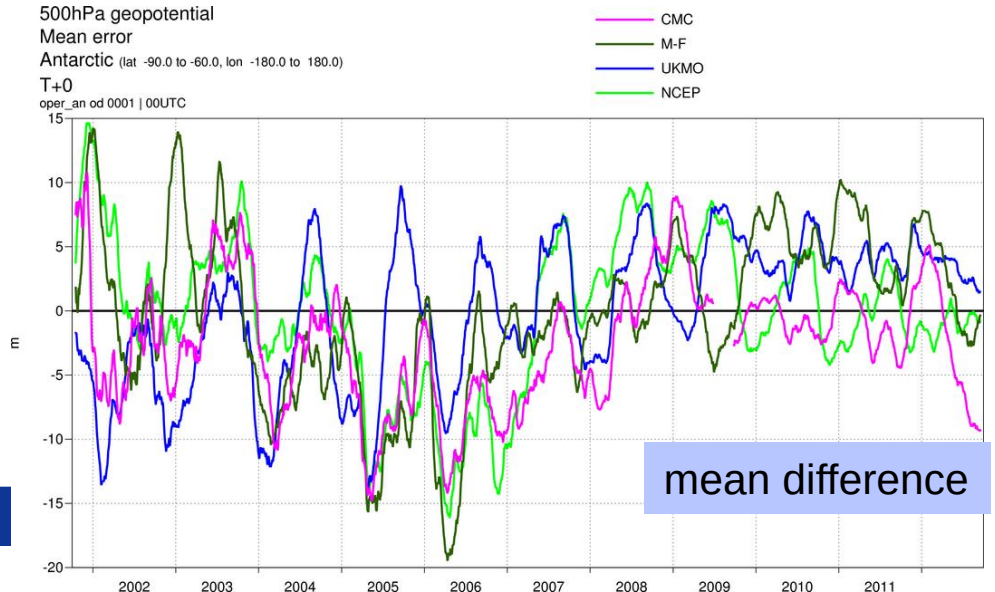




Comparison between analyses (Antarctic)

- Differences between the analyses of different centres
- Z500 30 day moving average
- Decrease over last decade in the difference between the analyses of different centres



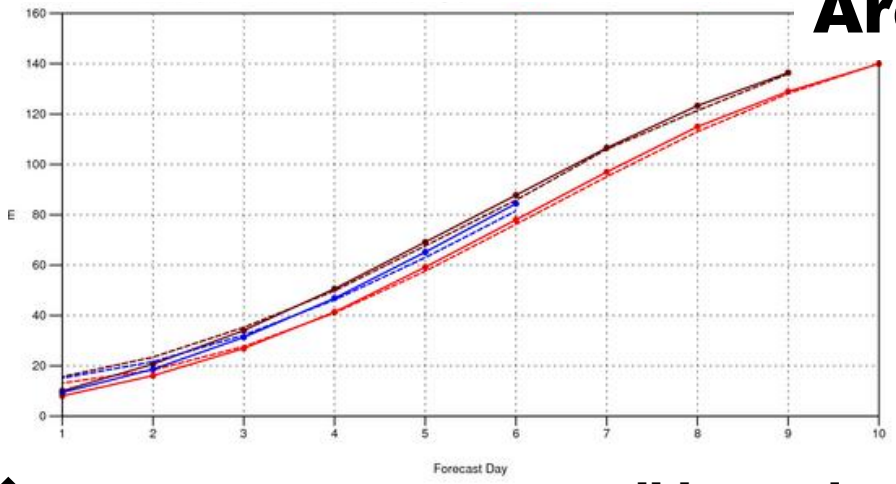


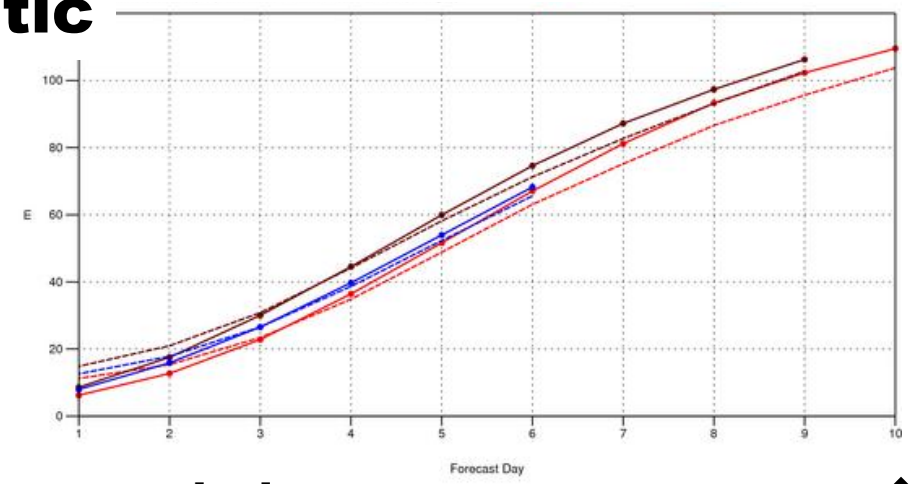
Seasons comparison

- JMA, MetOffice, ECMWF
- RMS error and bias against own analysis or radiosonde observations
- Arctic (top) vs Antarctic (bottom); December 2011 – February 2012 (left) vs June – August 2012 (right)

Seasons comparison – RMSE geopotential 500hPa

Arctic



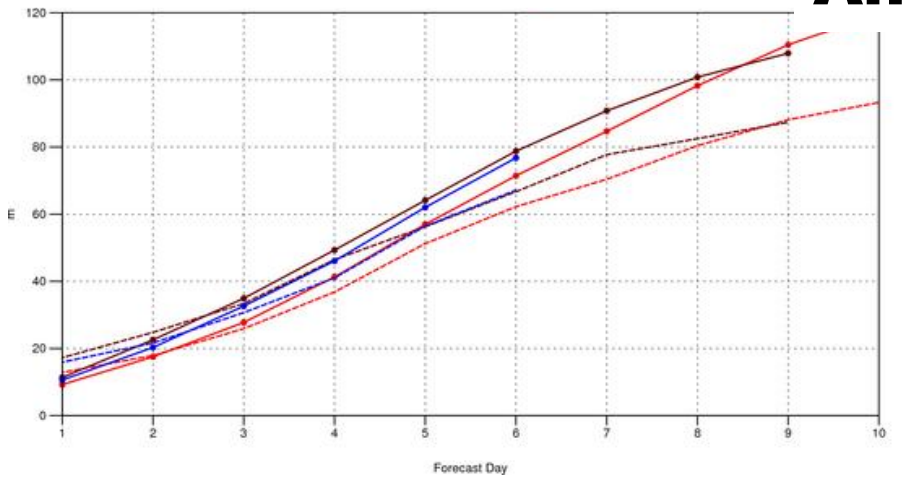


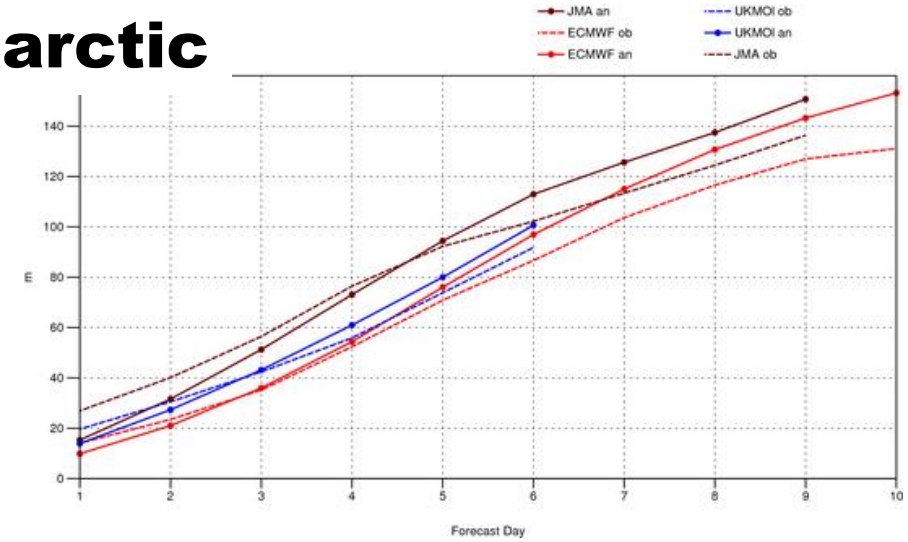
DJF 2012

solid – against own analysis
dashed – against observations

JJA 2012

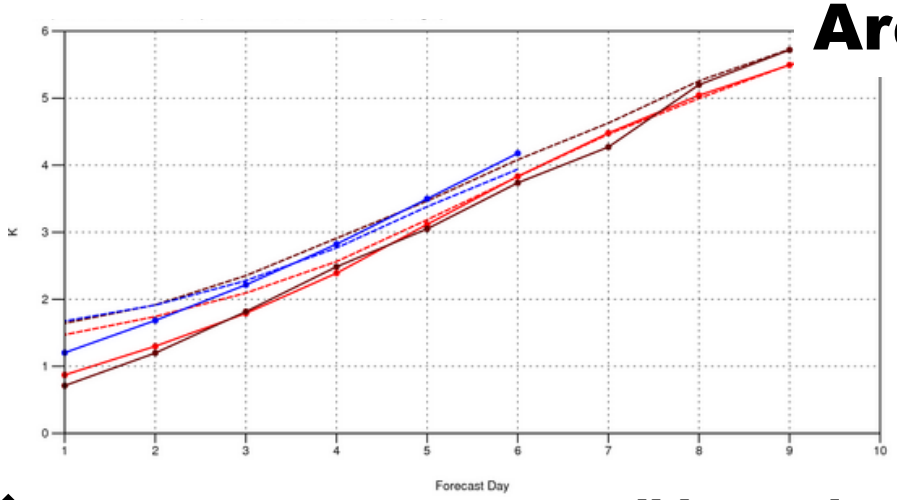
Antarctic

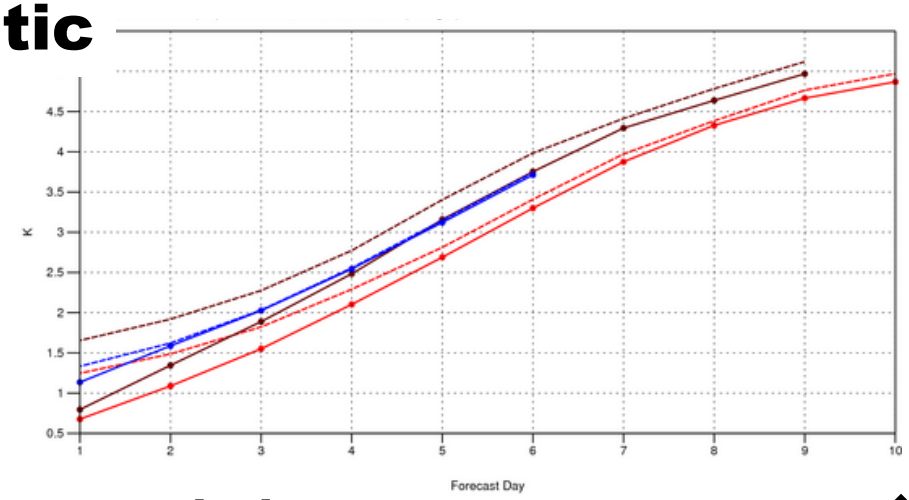




Seasons comparison – RMSE temperature 850hPa

Arctic



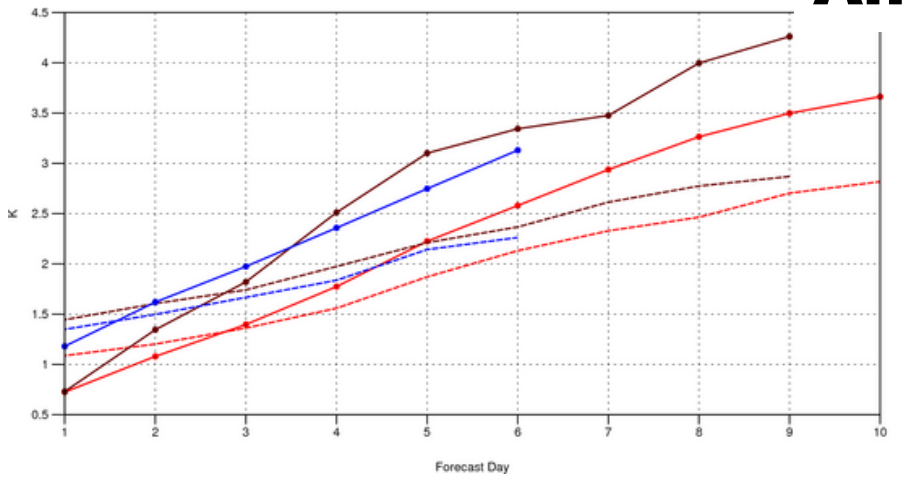


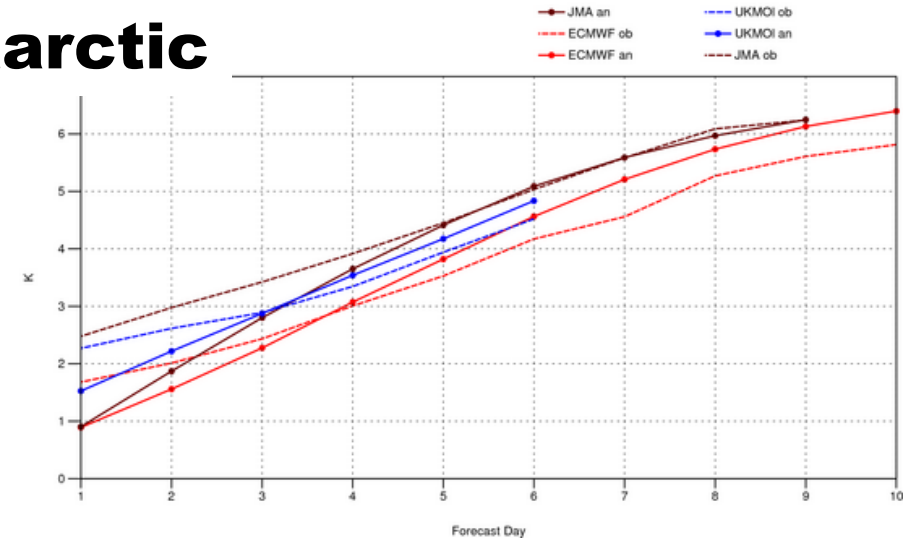
DJF 2012

solid – against own analysis
dashed – against observations

JJA 2012

Antarctic





Seasons comparison – bias temperature 850hPa

Arctic

DJF 2012

solid – against own analysis
dashed – against observations

JJA 2012

Antarctic

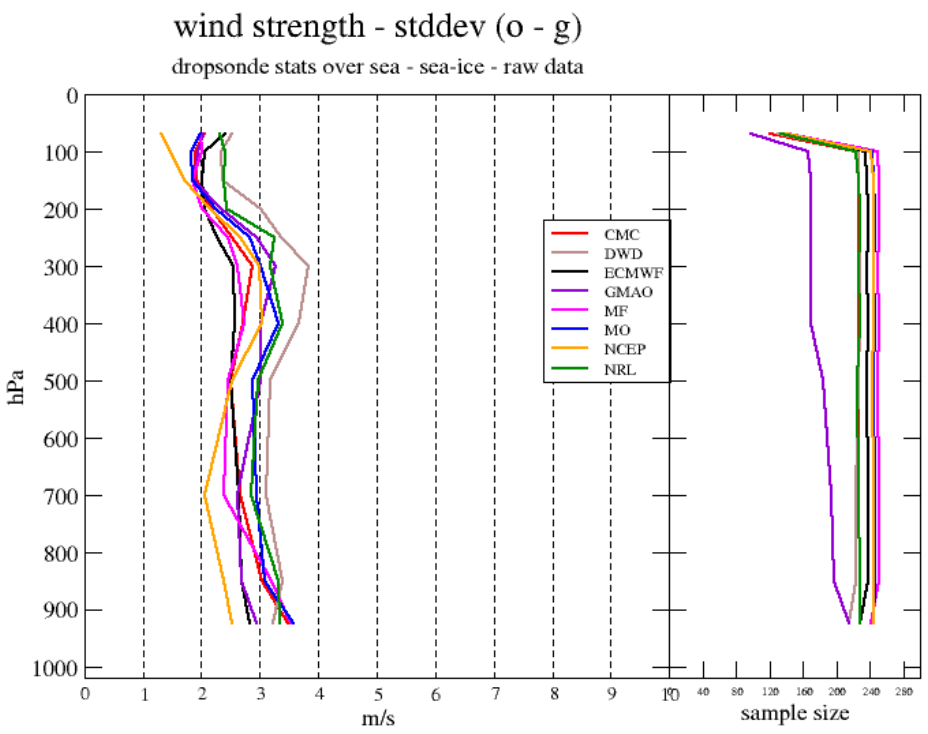
ECMWF

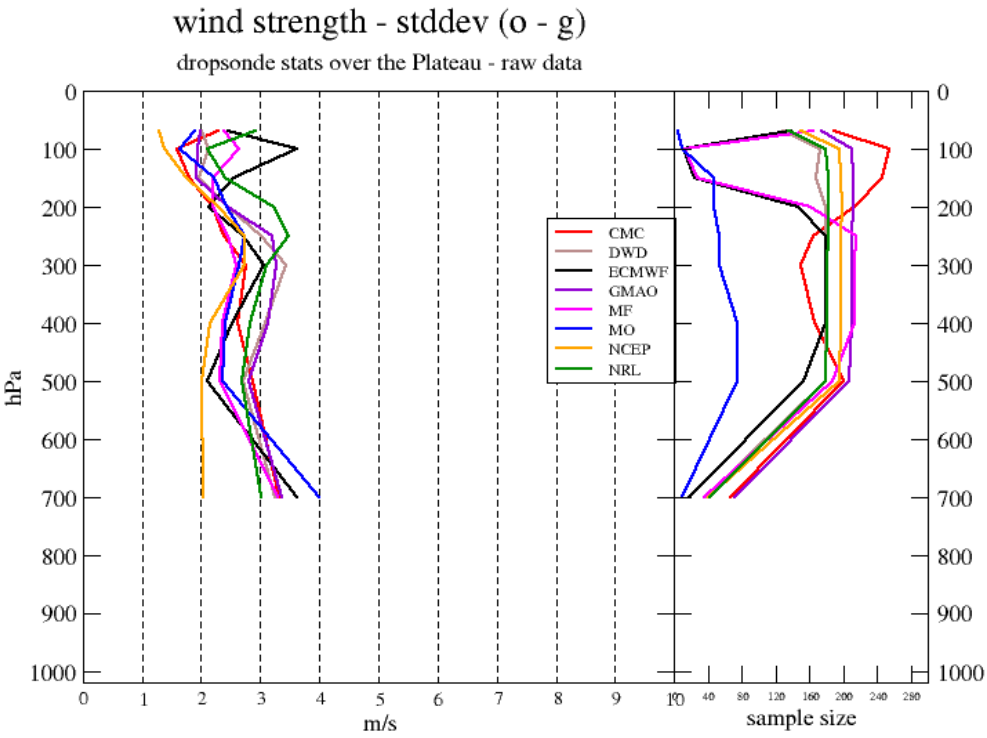
Seasons comparison: conclusions

- Errors larger in winter seasons
- Errors significantly larger in polar regions compared to extratropical hemispheric averages
- Larger differences between scores wrt obs and AN over Antarctica (pb of orography?)
- Errors with respect to observations dominate at short range but errors against analyses get mostly larger after Day 2 to 3
- Roughly similar relations for geopotential and temperature

Concordiasi field campaign: Statistics over all dropsondes assimilated in NRT

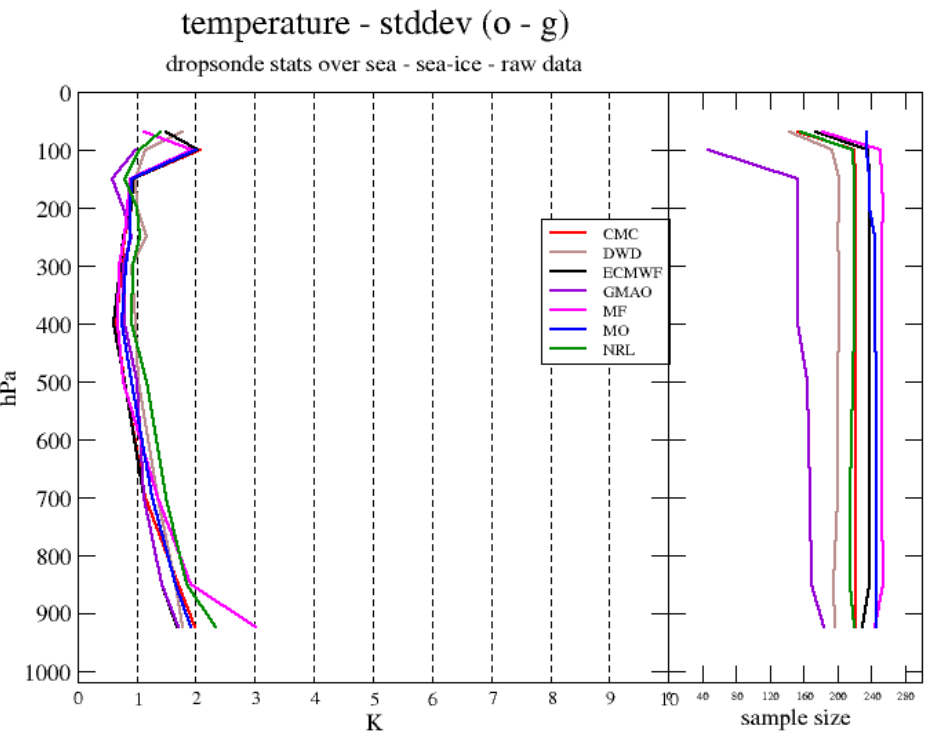
1 / 2 Wind speed

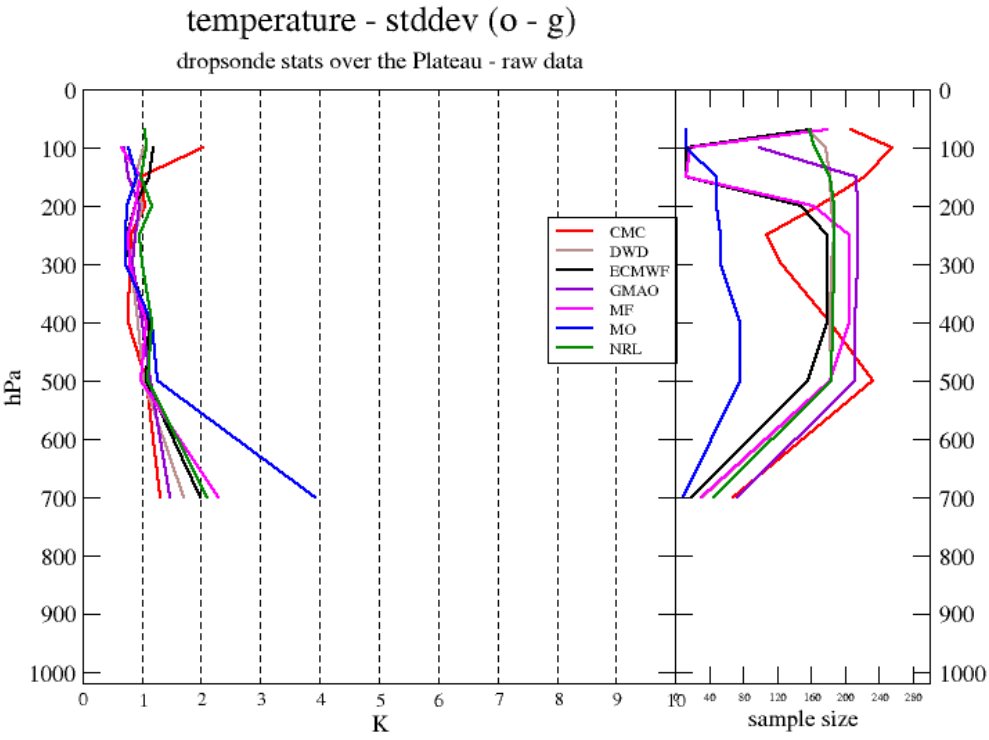




Concordiasi field campaign: Statistics over all dropsondes assimilated in NRT

2 / 2 Temperature

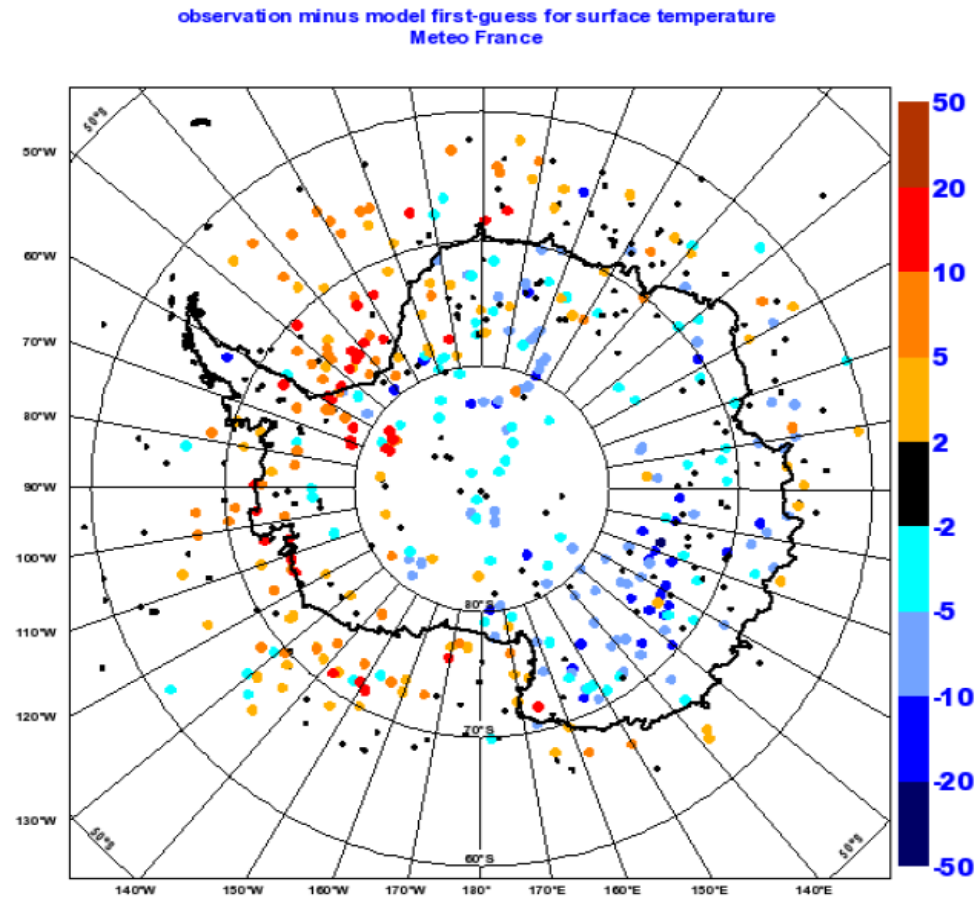


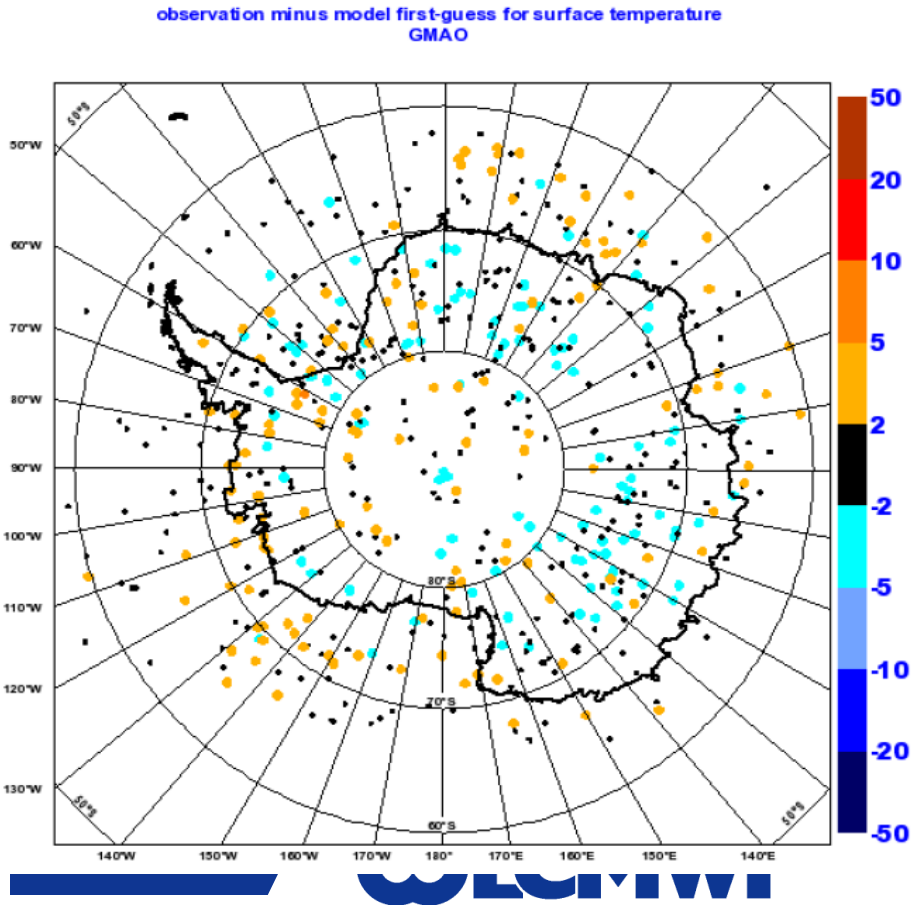


Which reality behind the statistics ?

Example: Obs *minus* model for T at the lowest dropsonde level provided by each centre

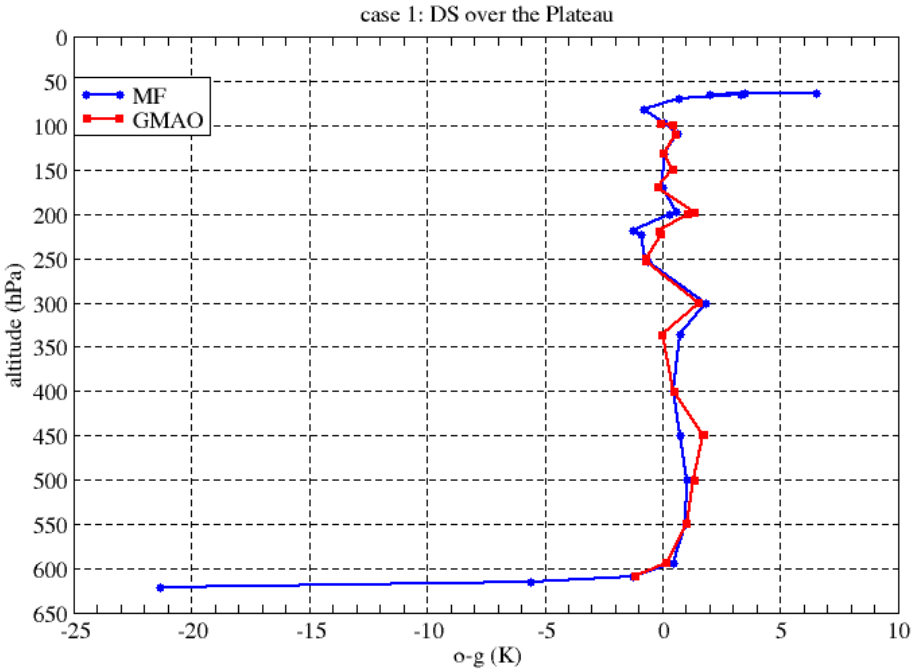
Is GMAO better than MF, smoother or more QC-ed ?

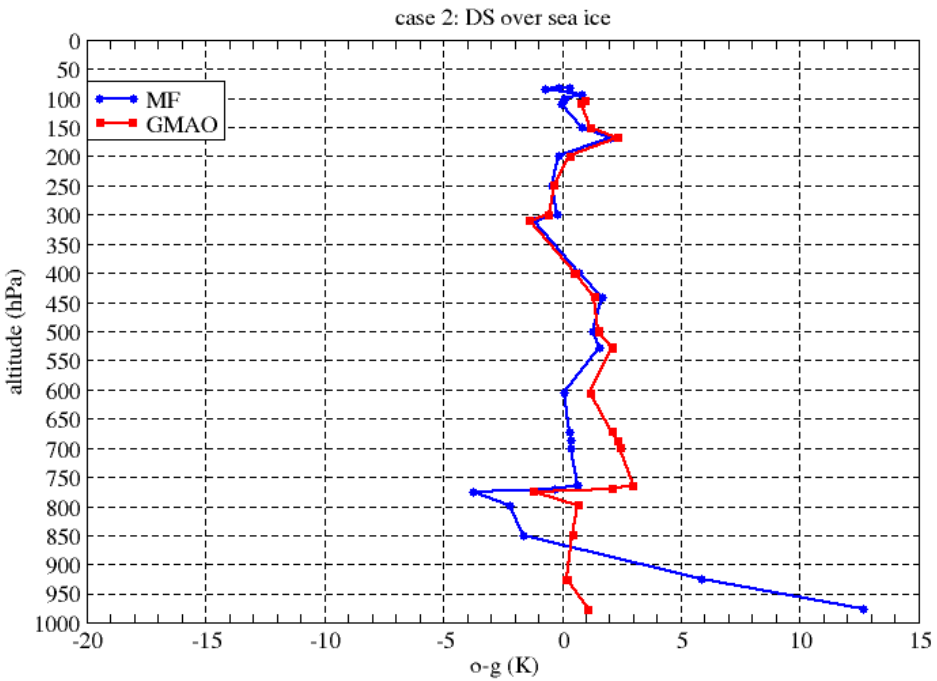




Which reality behind the statistics ?

Two profiles of obs *minus* model for T are selected





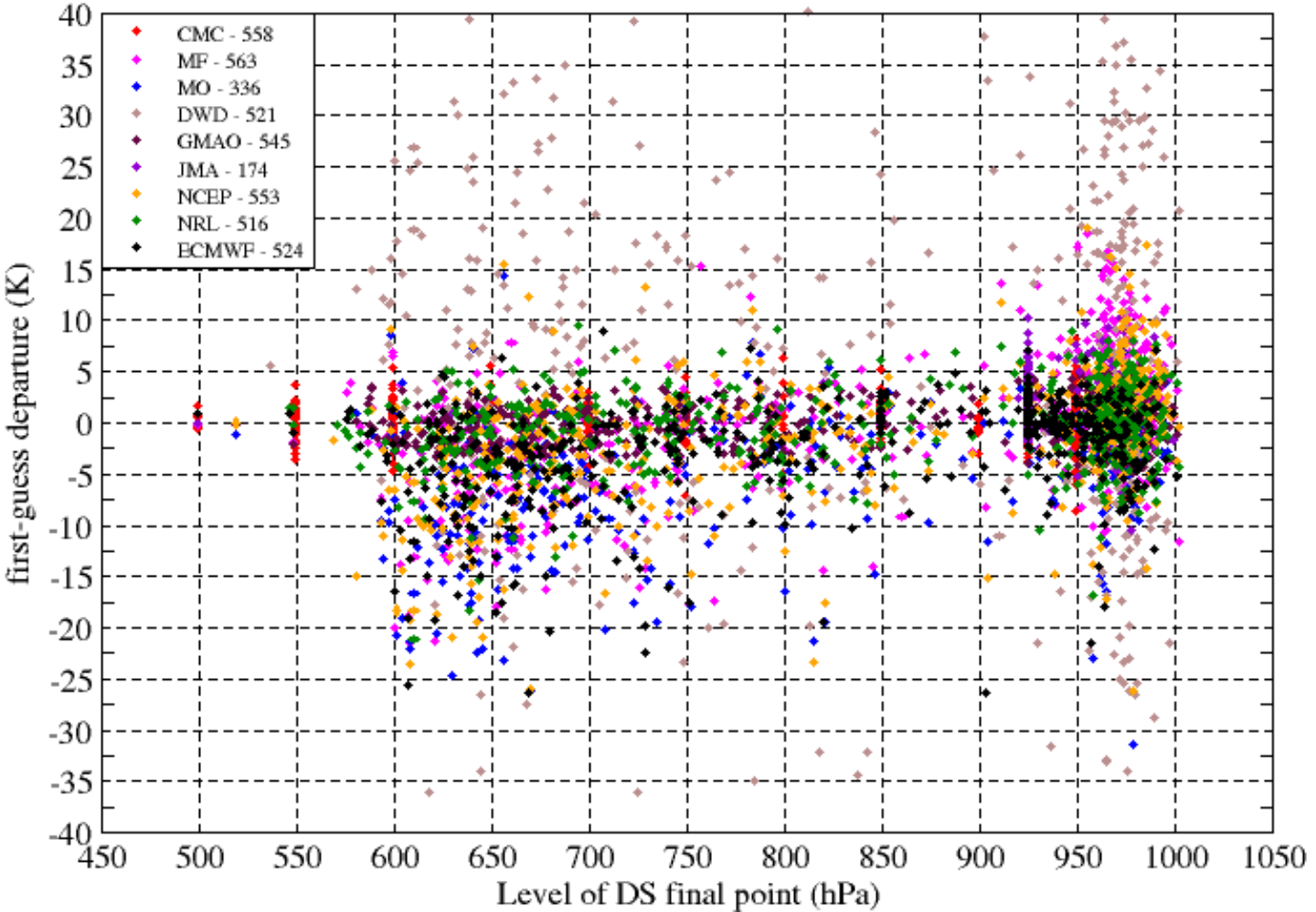
GMAO applied a QC, thus less gross errors occured near the surface over the Plateau

However, GMAO has a much better description of surface over sea-ice which implies better statistics

Obs *minus* model – T @ lowest DS level

All centres tend to be too warm over the Plateau.
Over sea-ice, no general pattern: some centres are too cold (eg. MF or NCEP), some centres sometimes too warm (eg. ECMWF or MO), and some centres seem to be OK

Caution !
Some centres provided Qced statistics (with various thresholds)
While other centres provided all statistics



Conclusions

- General scores:
  - Versus own analyses
    - I forgot which Centre is the red curve?
    - NCEP recent improvements noticeable
    - BOM and KMA new systems
  - Versus observations
    - Differences get smaller and smaller
- Please have a look at http://apps.ecmwf.int/wmolcdnv/
- Progress in polar verification
  - Verif against observations
  - Still some way to go – we need more participants!
  - Concordiasi observational dataset as independent verification.