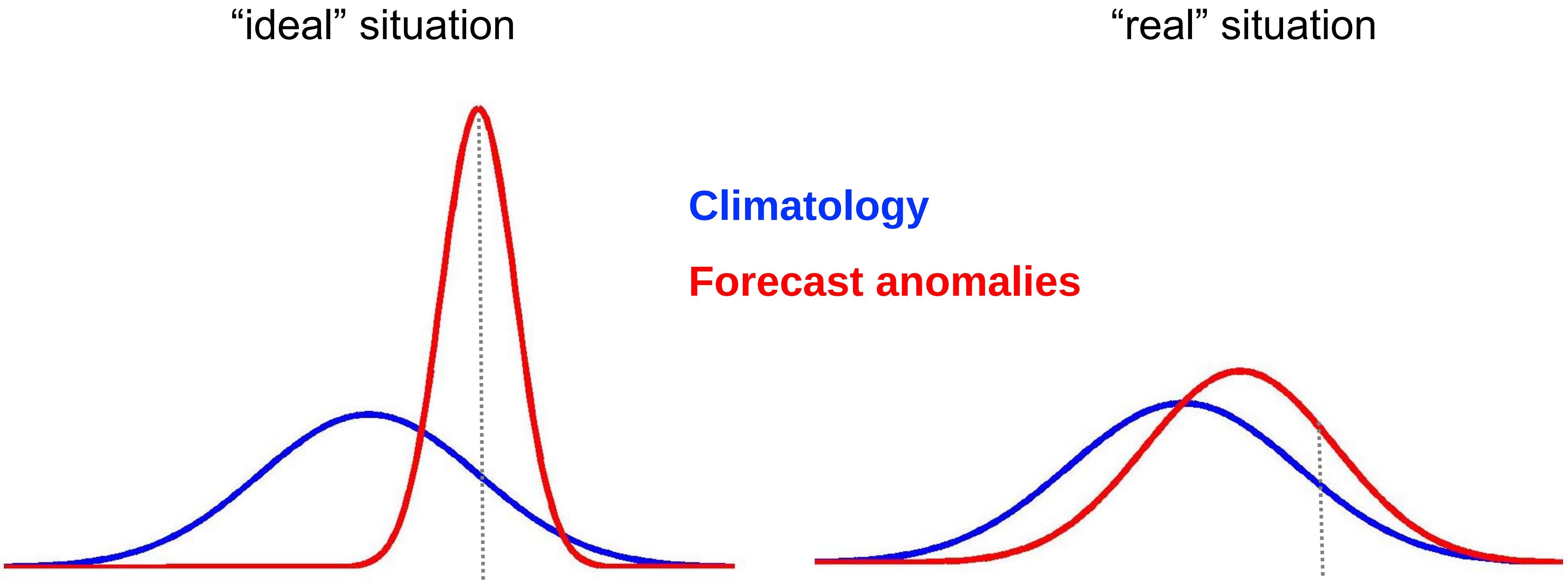
Seasonal forecast ensembles:

How confident are predictability estimates of the NAO?

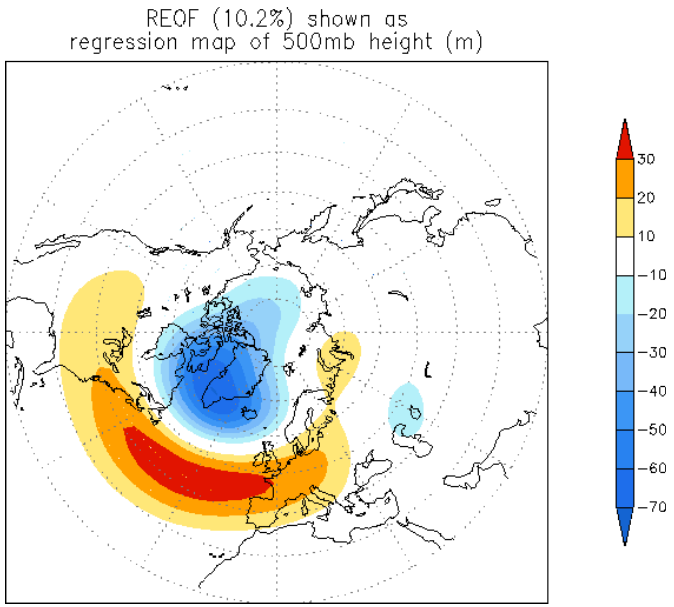
Antje Weisheimer

ECMWF
University of Oxford

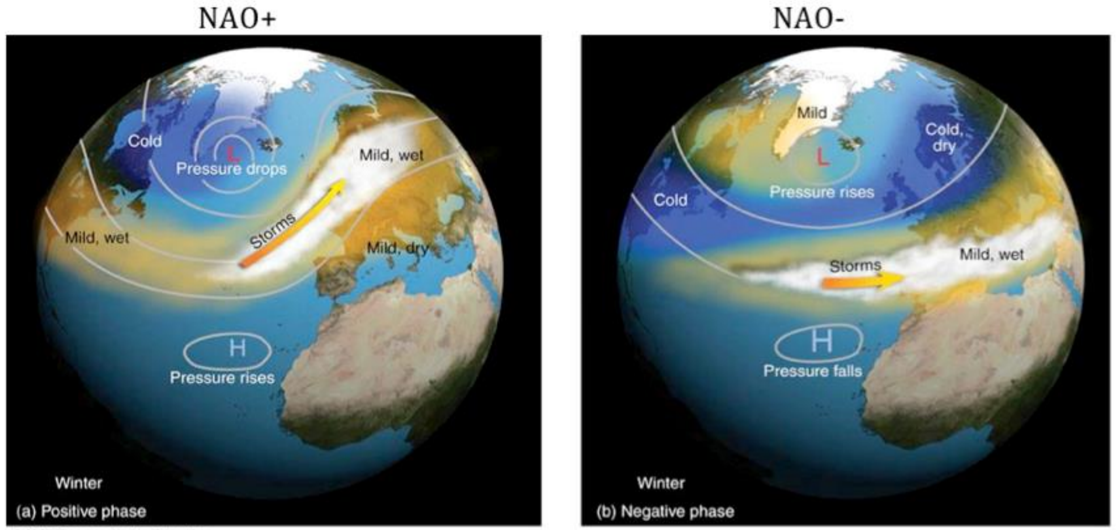
Seasonal forecasting in the extra-tropics
with low signal-to-noise ratios



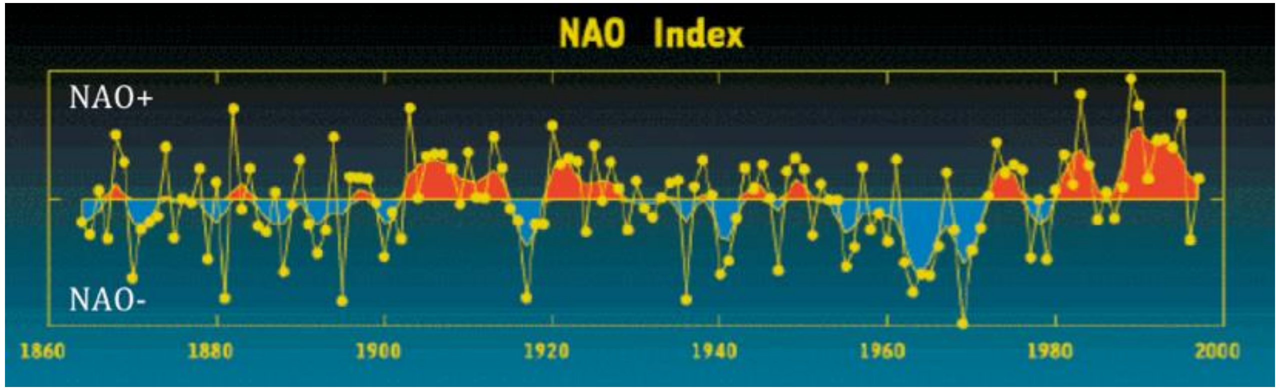
The North Atlantic Oscillation (NAO)



- Dominant mode of variability on a range of time scales over the North Atlantic-European region
- Typically defined as the 1st EOF of MSLP or Z500
- NAO index: 1st Principal Component or sometimes (mostly for historical reasons) as normalised MSLP difference between Iceland and the Azores

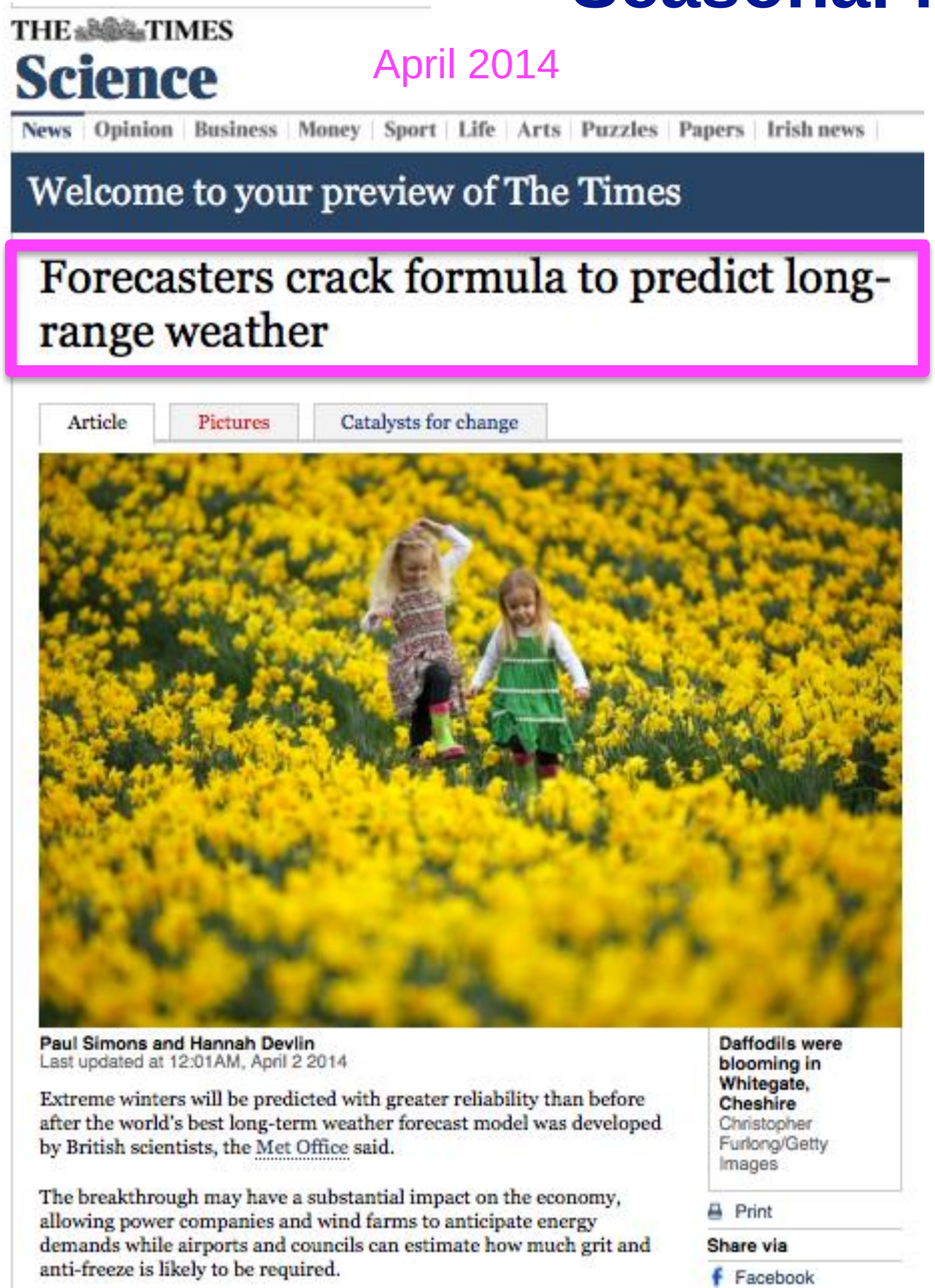


© 2007 Thomson Higher Education



Seasonal forecasts of the winter NAO

April 2014



THE TIMES

Science

News | Opinion | Business | Money | Sport | Life | Arts | Puzzles | Papers | Irish news |

Welcome to your preview of The Times

Forecasters crack formula to predict long-range weather

Article | Pictures | Catalysts for change

Daffodils were blooming in Whitegate, Cheshire
Christopher Furlong/Getty Images

Paul Simons and Hannah Devlin
Last updated at 12:01AM, April 2 2014

Extreme winters will be predicted with greater reliability than before after the world's best long-term weather forecast model was developed by British scientists, the Met Office said.

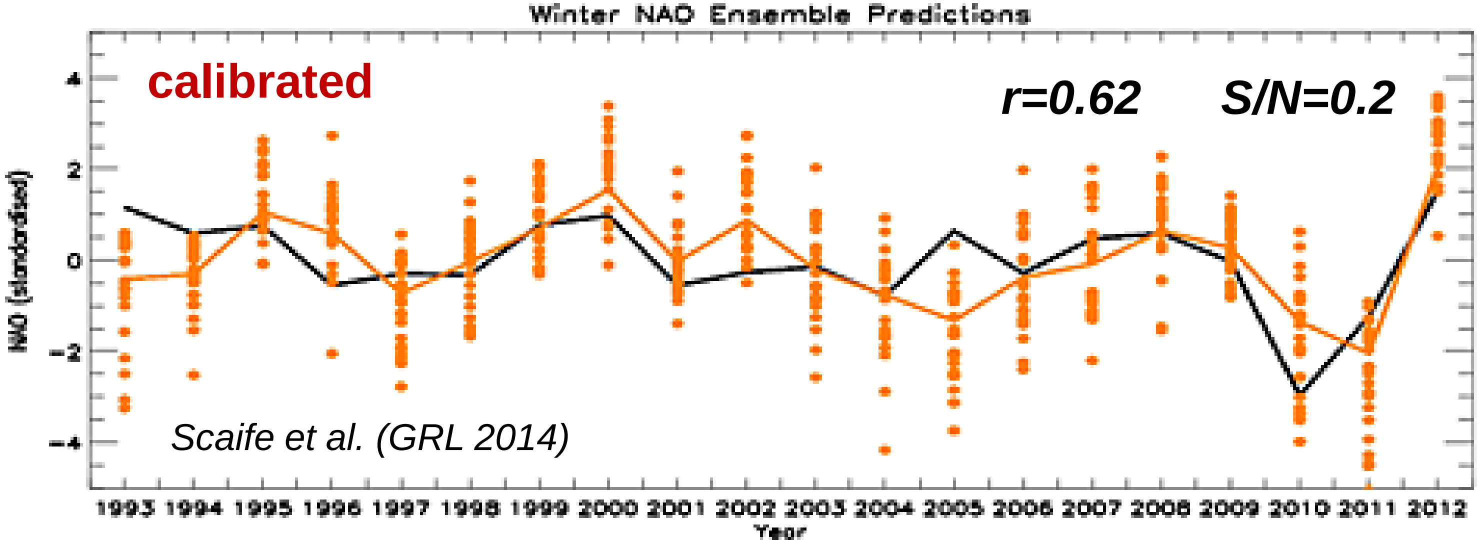
The breakthrough may have a substantial impact on the economy, allowing power companies and wind farms to anticipate energy demands while airports and councils can estimate how much grit and anti-freeze is likely to be required.

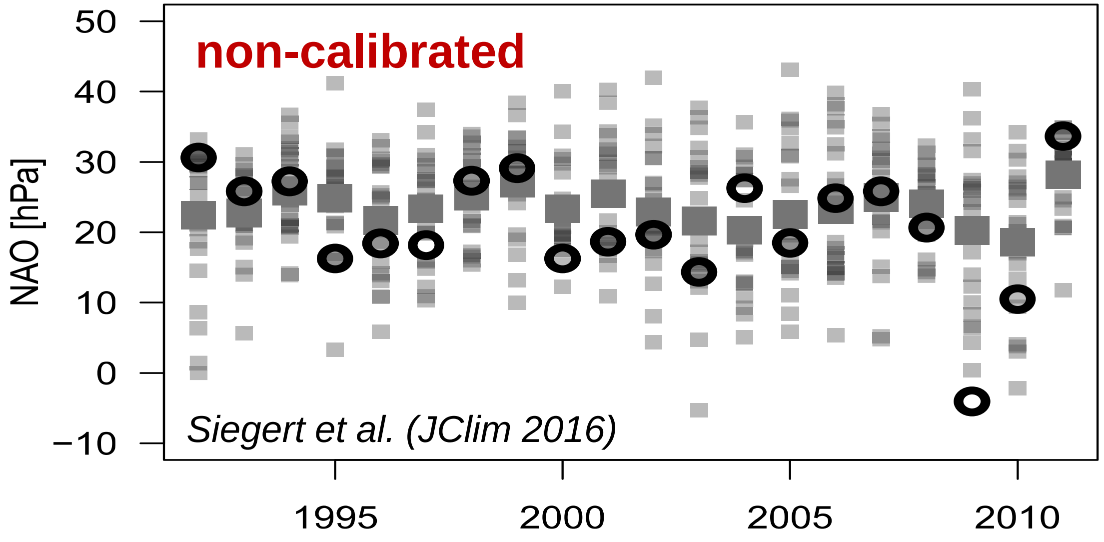
Print

Share via

Facebook

Ensemble hindcasts of the NAO index 1993-2012
with the Met Office model (GloSea GA3)





Signal and noise

$x_{m,n}$: *variable x with member m and year n*

Mean: $\bar{x} = \frac{1}{NM}\sum_{n}^{N}\sum_{m}^{M} x_{m,n}$

Ensemble mean: $\langle x_n \rangle = \frac{1}{M}\sum_{m}^{M} x_{m,n}$

Variance: $VAR_{total} = \frac{1}{NM}\sum_{n}^{N}\sum_{m}^{M}\left(x_{m,n} - \bar{x}\right)^2$

$$\boldsymbol{VAR_{total} = VAR_{signal} + VAR_{noise}}$$

→ ***S/N = VAR$_{signal}$ / VAR$_{noise}$***

$$VAR_{signal} = \frac{1}{N}\sum_{n}^{N}\left(\langle x_n \rangle - \bar{x}\right)^2$$

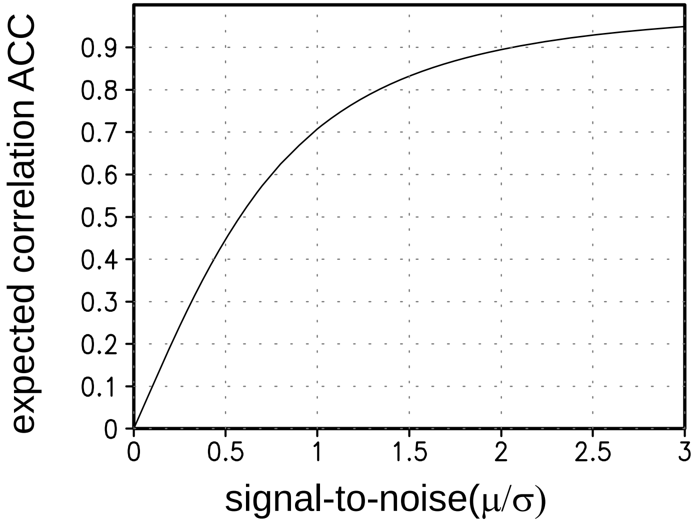
ensemble mean variance
*→ **"signal"***

$$VAR_{noise} = \frac{1}{NM}\sum_{n}^{N}\sum_{m}^{M}\left(x_{m,n} - \langle x_n \rangle\right)^2$$

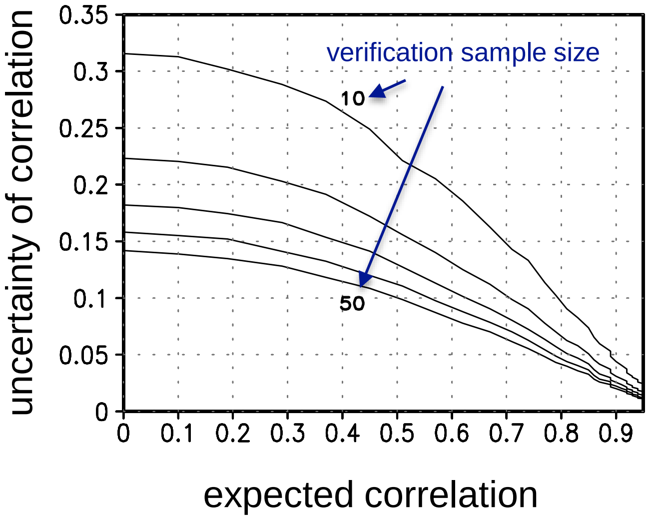
*variance of ensemble members about ensemble mean (=spread) → **"noise"***

Correlation skill and signal-to-noise (S/N) ratio

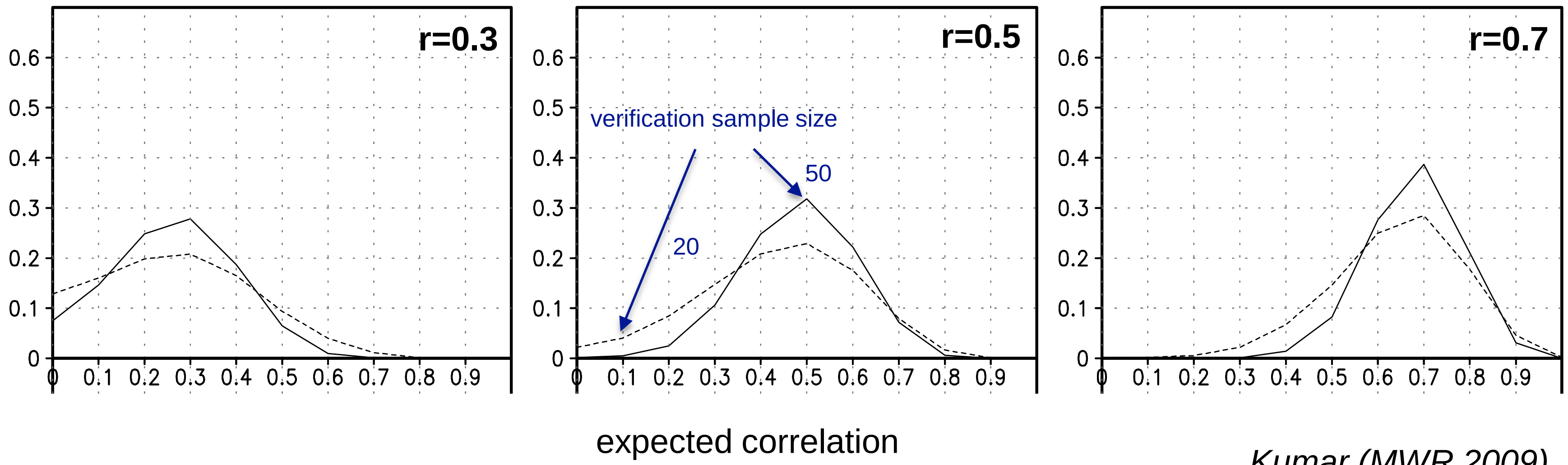
The *expected value* for various measures of skill for seasonal climate predictions is determined by the S/N ratio.



"The expected value, however, is only realized for long verification time series. In practice, the verifications for specific seasons seldom exceed a sample size of 30. The estimates of skill measure based on small verification time series, because of sampling errors, can have large departures from their expected value."



Probability of expected correlation for a given realised value of skill



Kumar (MWR 2009)

The Ratio of Predictable Components (RPC)

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2014GL061146

Key Points:
- Model members can be too noisy and not potential realizations of the real world
- Predictability may be underestimated by idealized experiments and skill measures
- Can achieve skilful and reliable forecasts using large ensembles to reduce noise

Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?

Rosie Eade[1], Doug Smith[1], Adam Scaife[1], Emily Wallace[1], Nick Dunstone[1], Leon Hermanson[1], and Niall Robinson[1]

[1]Met Office Hadley Centre, Exeter, UK

Abstract Seasonal-to-decadal predictions are inevitably uncertain, depending on the size of the predictable

$$RPC = \frac{PC_{obs}}{PC_{model}} \geq \frac{r(obs, ens\ mean)}{\sqrt{VAR_{signal}/VAR_{total}}}$$

Predictable Components (PCs) … predictable part of the total variance

observed Pc_{obs} … estimated from explained variance = ***r*²*(obs, ensmean)***

model Pc_{model} … estimated from ratio of signal variance to total variance

Eade et al. (GRL 2014)

Perfect model ensembles and potential skill

What is a perfect model ensemble?

- Perfect sampling of the underlying probability distribution of the true state
- Over a large number of forecasts, the statistical properties of the truth are identical to the statistical properties of a member of the ensemble
- I.e., the truth is indistinguishable from the ensemble

→ Replace observation with ensemble member

Perfect model ensembles and potential skill

Properties of a perfect model ensemble

- Time-mean ensemble spread == RMSE of ensemble mean forecast
- *r* (perfect model) = corr(ens mean,ens members) → "*potential skill*"
- RPC of a perfect ensemble == 1
- Observed correlation ≤ perfect model correlation ??

Perfect model ensembles and potential skill

Implications for non-perfect ensembles

- Time-mean ensemble spread ≠ RMSE of ensemble mean forecast

  ensemble spread < RMSE → ensemble is *underdispersive*

  ensemble spread > RMSE → ensemble is *overdispersive*

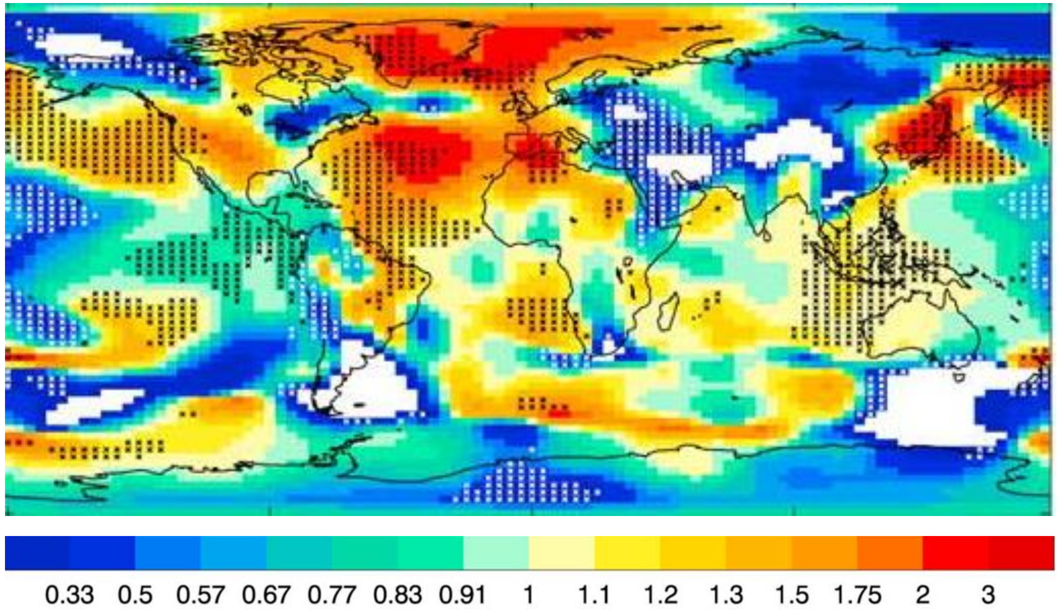
- RPC ≠ 1

  RPC > 1 → underconfidence; VAR_{signal} too small, model underestimates predictability of real world, observed correlation > perfect model correlation

  RPC < 1 → overconfidence; observed correlation < perfect model correlation model predictability is larger than in real world

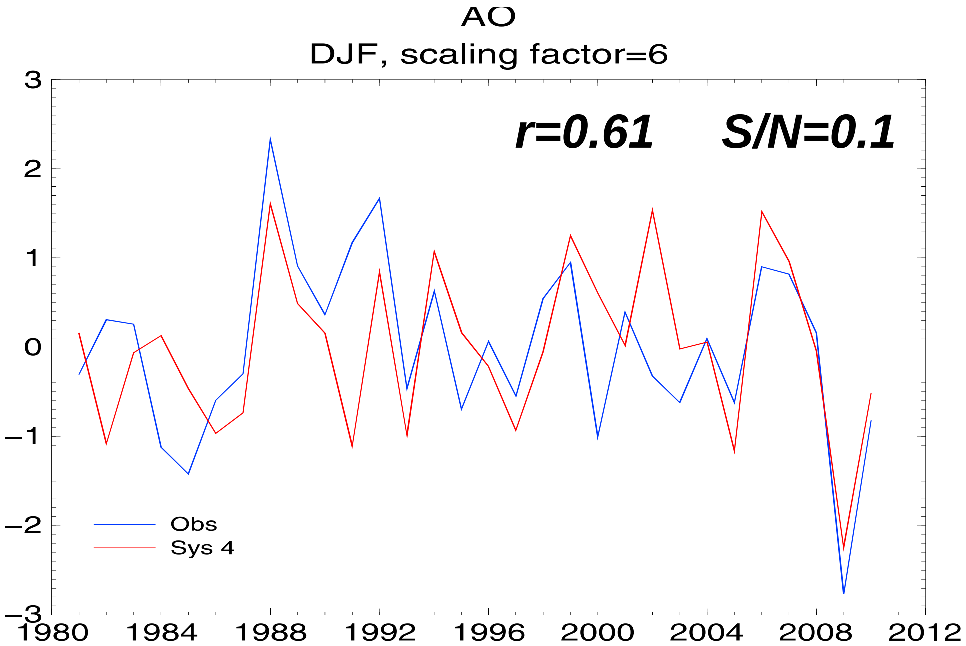
The signal-to-noise “conundrum” or “paradox”

RPC of DJF MSLP in GloSea5 $\left(\mathrm{RPC} = \frac{\mathrm{PC_{obs}}}{\mathrm{PC_{model}}}\right)$



Eade et al. (GRL 2014)

Ensemble hindcasts of the AO index 1981-2010 with the ECMWF System 4



Stockdale et al. (GRL 2015)

The real world seems to have higher predictability than the model.

The signal-to-noise "conundrum" or "paradox"

A Bayesian Framework for Verification and Recalibration of Ensemble Forecasts: How Uncertain is NAO Predictability?

STEFAN SIEGERT, DAVID B. STEPHENSON, AND PHILIP G. SANSOM
University of Exeter, Exeter, United Kingdom

ADAM A. SCAIFE, ROSIE EADE, AND ALBERTO ARRIBAS
Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received 11 March 2015, in final form 30 July 2015)

ABSTRACT

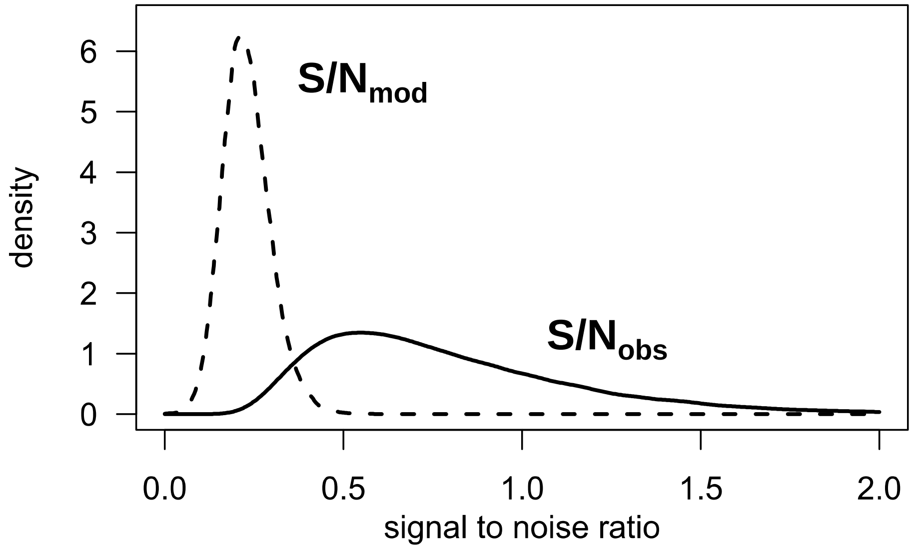
Predictability estimates of ensemble prediction systems are uncertain because of limited numbers of past

- 95% uncertainty intervals on *r=0.62* are [0.19;0.68]
- S/N_{obs} is larger than S/N_{model}

  → raw forecasts should not be taken as representative scenarios of the observations (not exchangeable)

  → predictable signal in model too weak
- The particular 20-yr period is unusual and produces higher-than-normal correlation skill

posterior distributions



Siegert et al. (JClim 2016)

What is the empirical evidence on shorter forecast ranges that

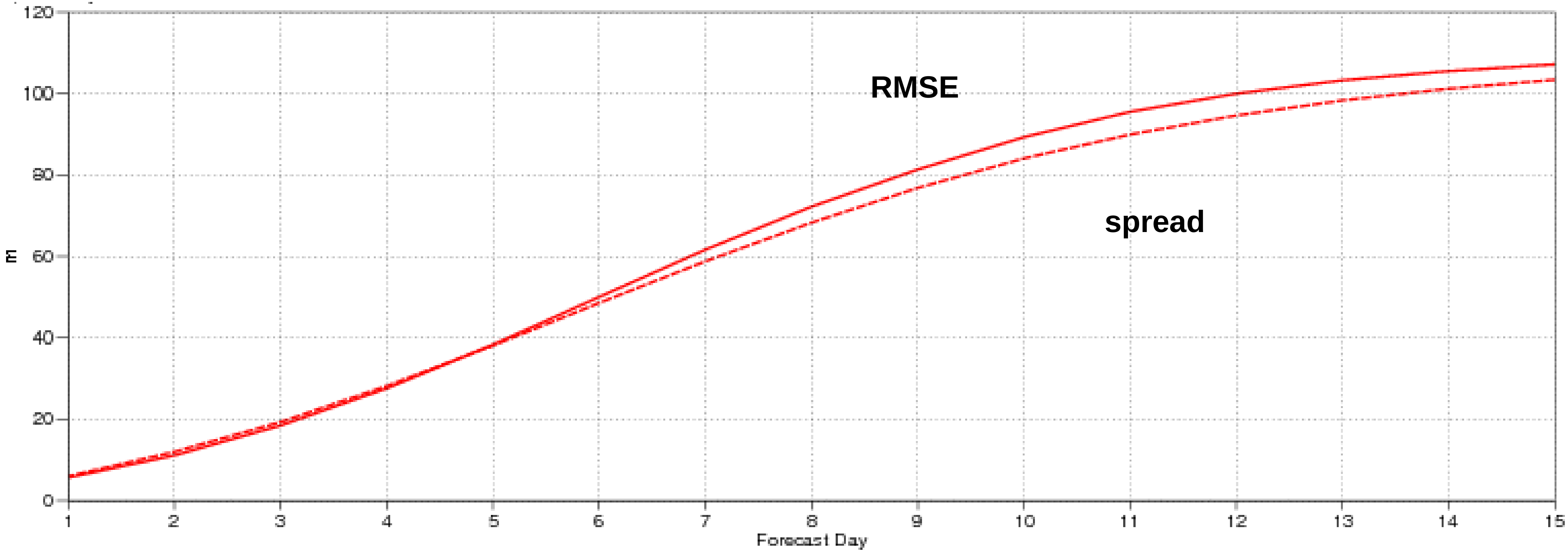
i) **models are overdispersive**

and/or

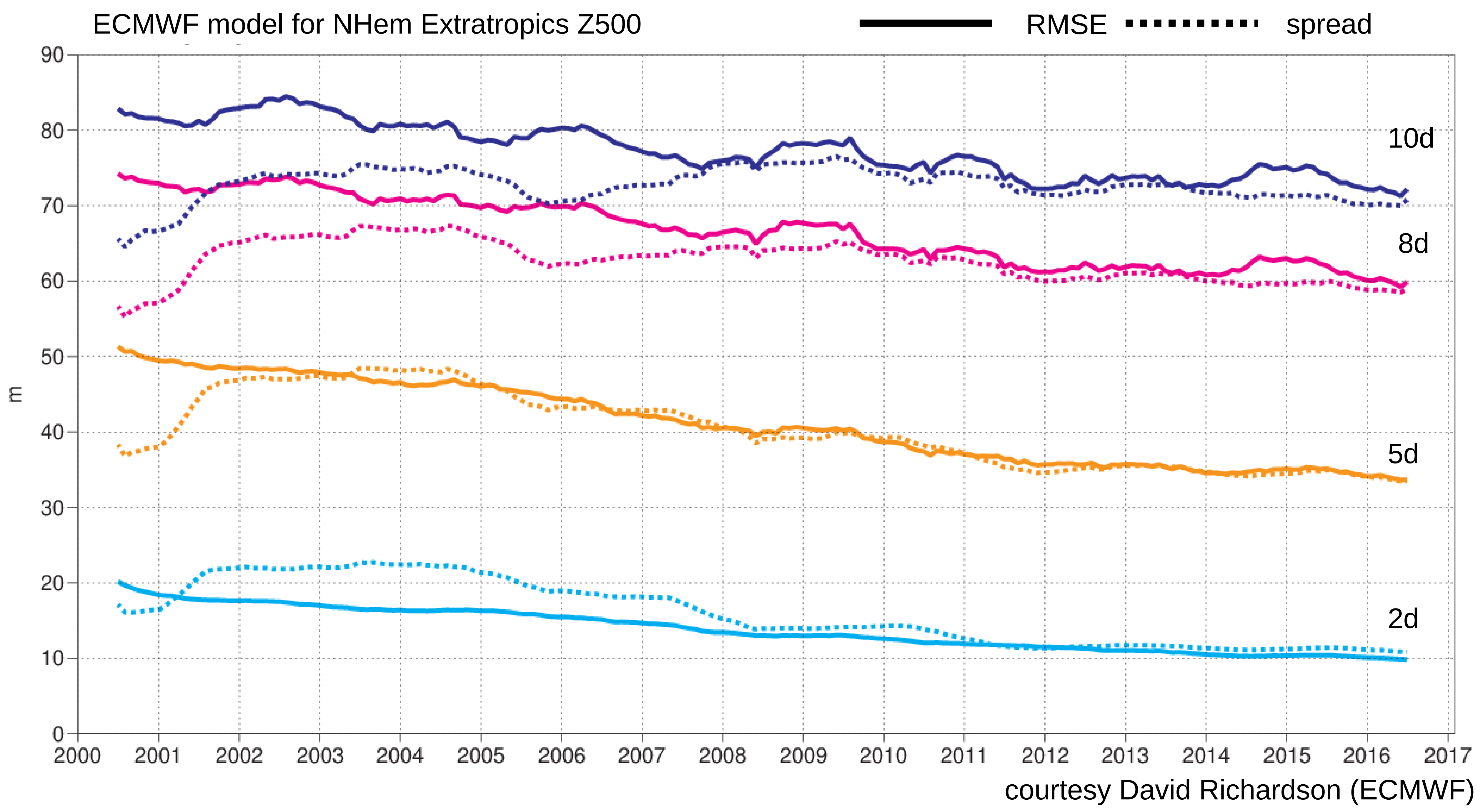
ii) **model estimates of predictability are too low (underconfidence)?**

Spread-skill relationship in medium-range forecasts

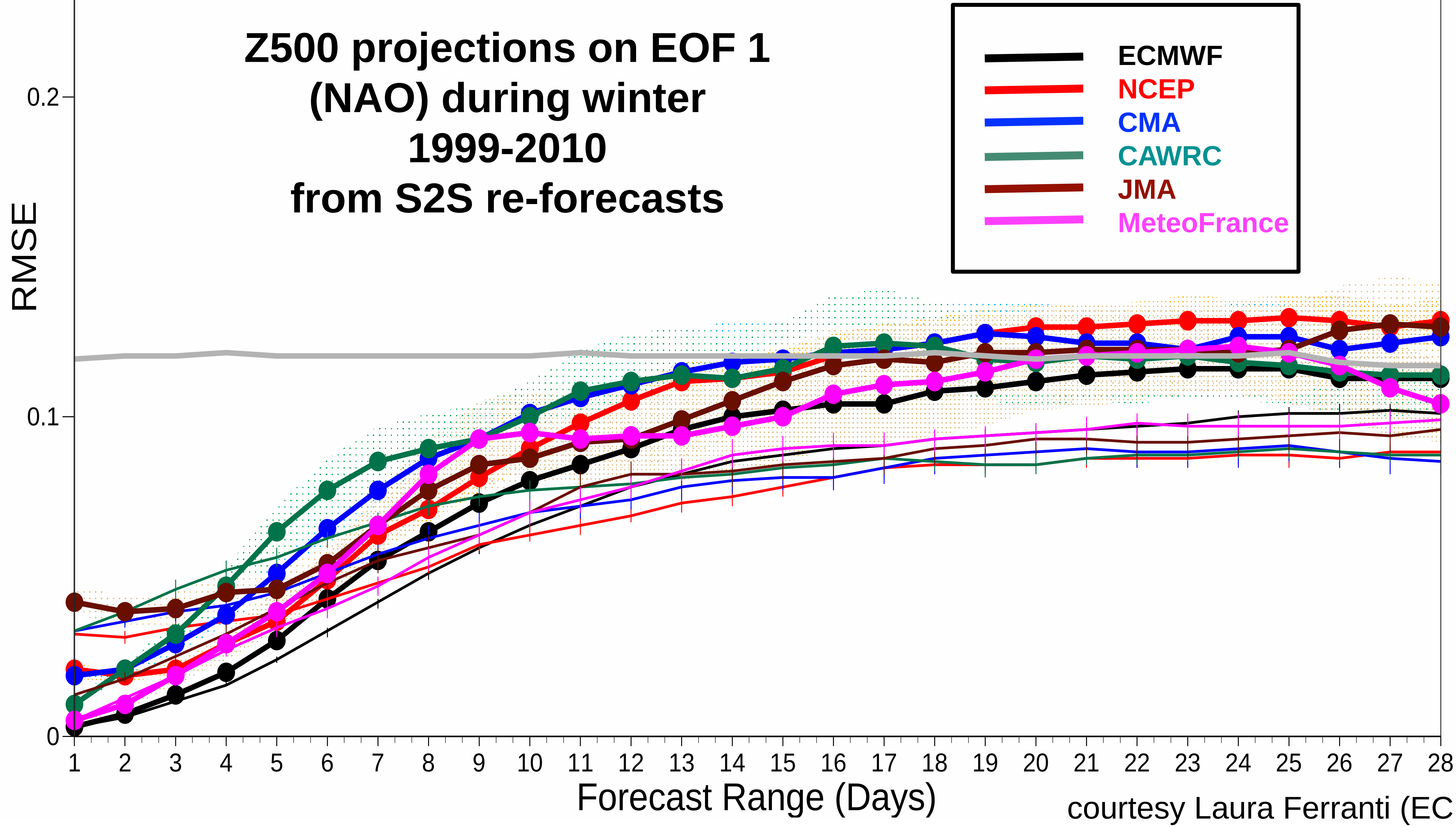
ECMWF model for NHem Extratropics Z500 DJF 2016/17



courtesy David Richardson (ECMWF)

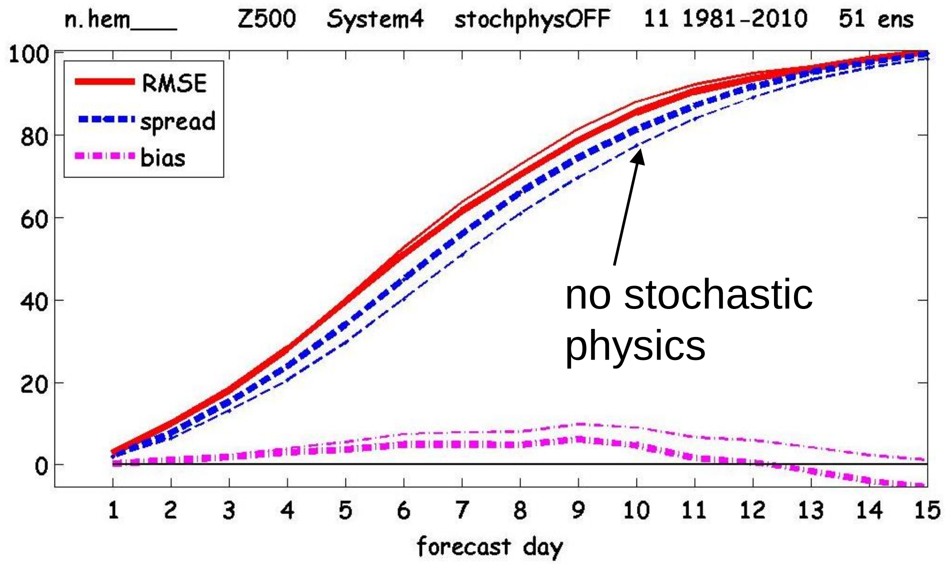

ECMWF model for NHem Extratropics Z500
RMSE
spread
90
80
70
60
50
40
30
20
10
0
m
10d
8d
5d
2d
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
courtesy David Richardson (ECMWF)

Spread-skill relationship in monthly forecasts



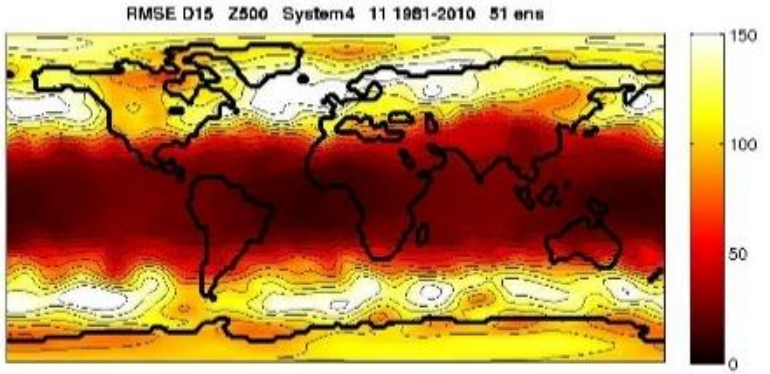
courtesy Laura Ferranti (ECMWF)

1st Nov start date 1981-2010 Z500 seasonal forecasts S4 51 ens members

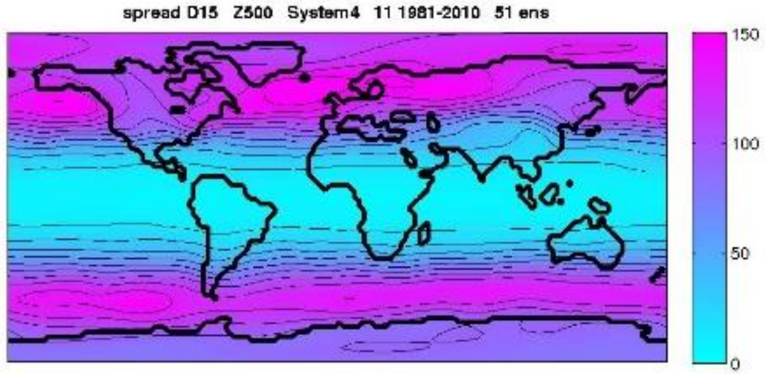


D+15

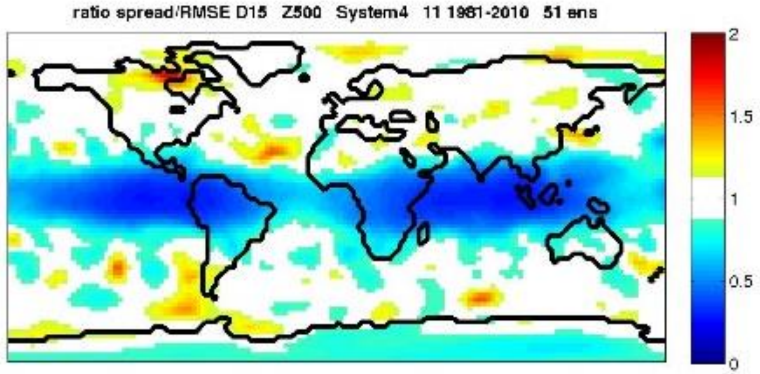
RMSE



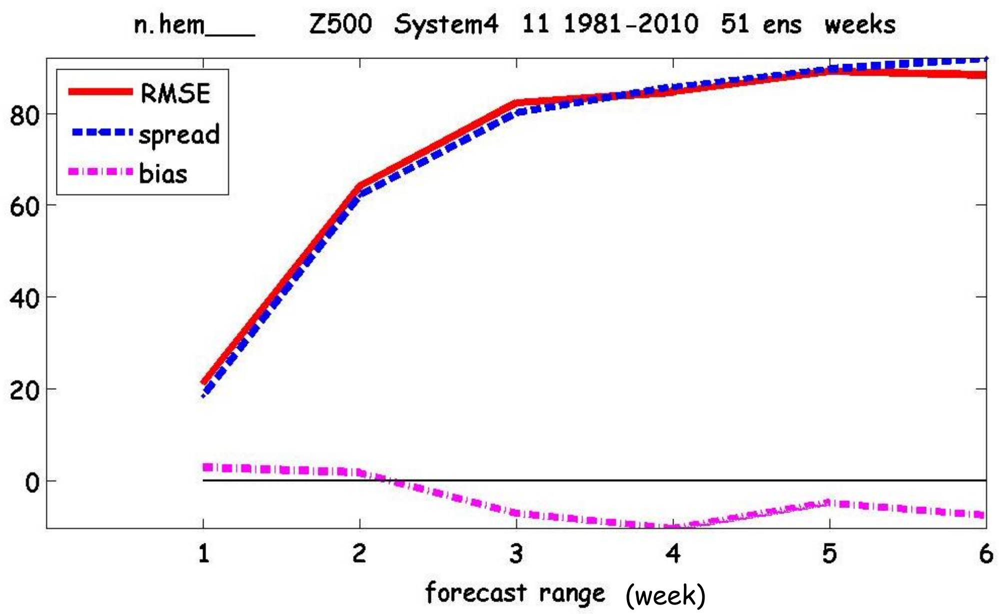
spread



spread/RMSE

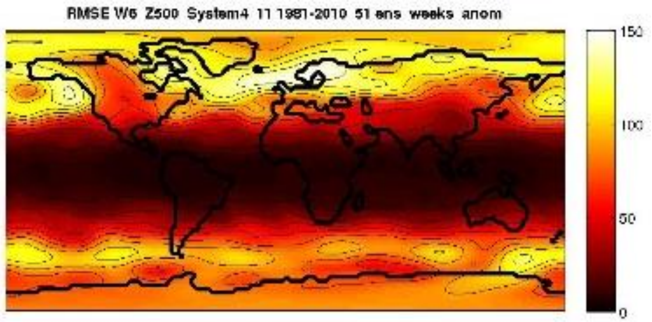


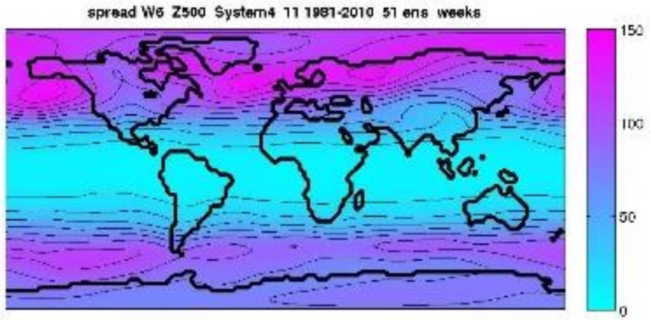
1st Nov start date 1981-2010 Z500 seasonal forecasts S4 51 ens members

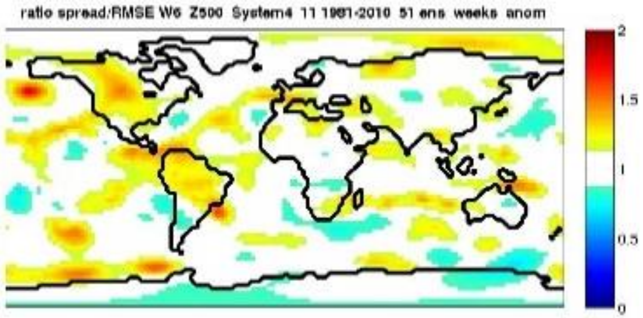


Week 6

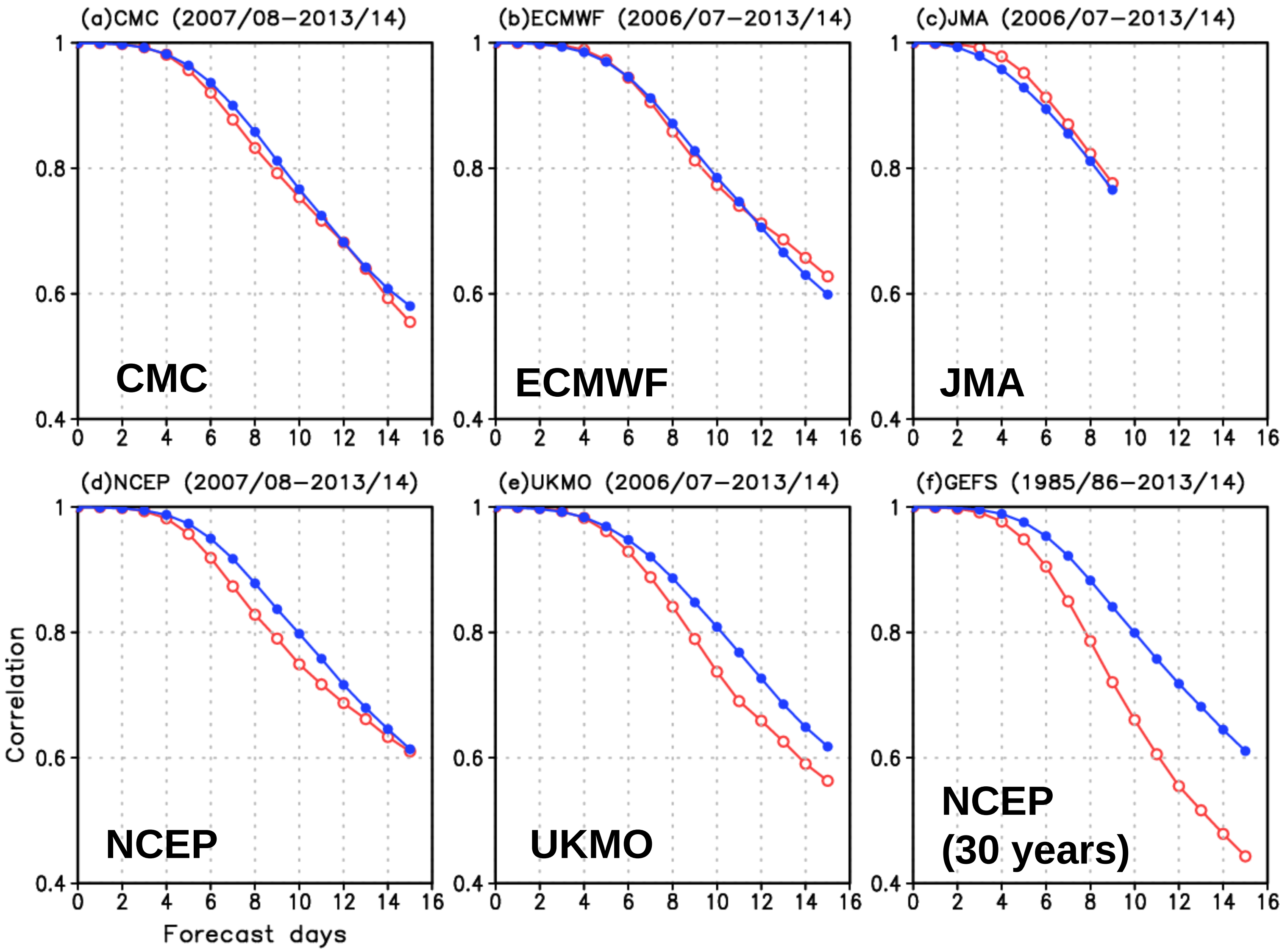
RMSE	spread	spread/RMSE







Correlations in medium-range forecasts (TIGGE models)



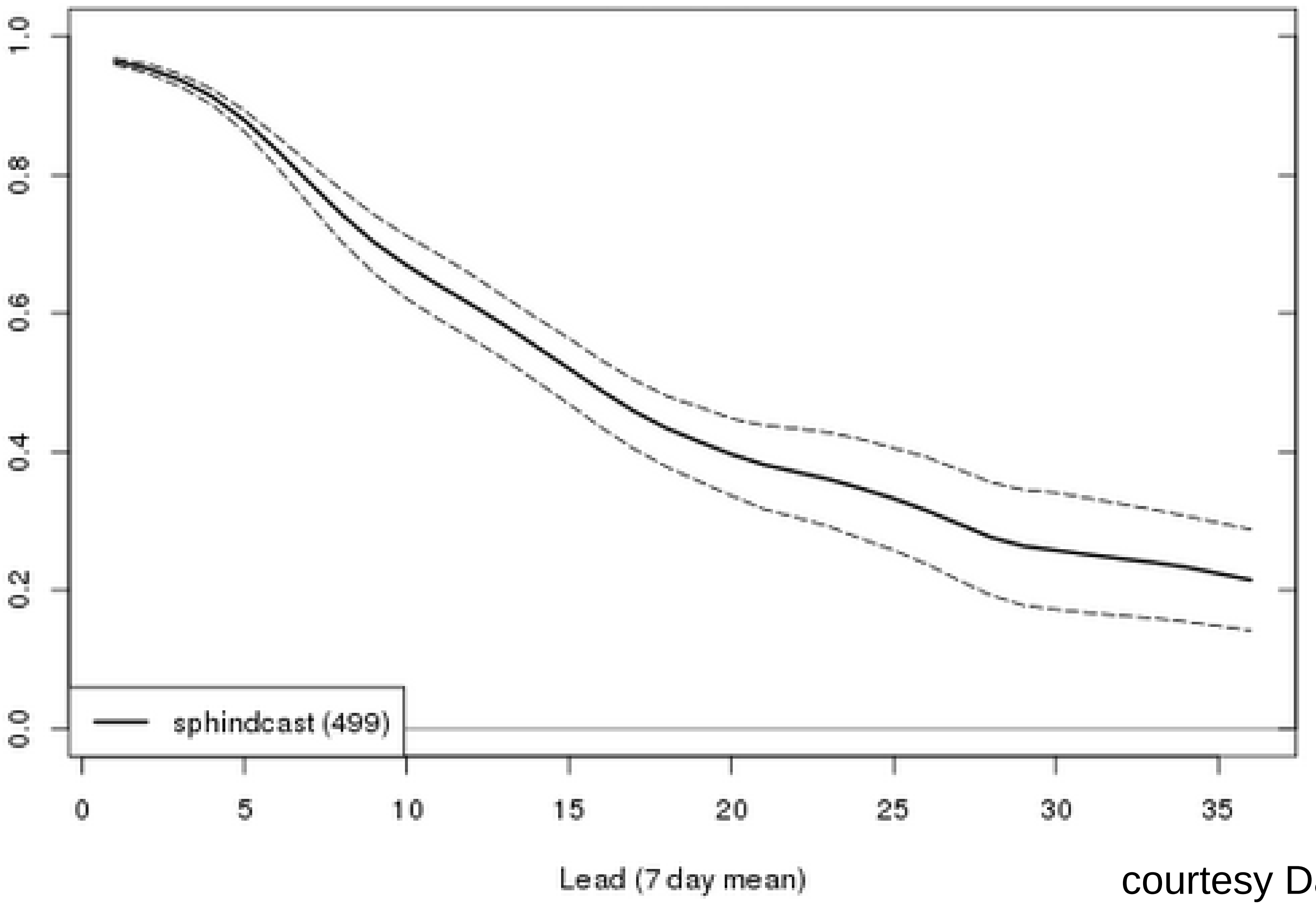
real world

model world
(potential or
perfect model skill)

courtesy Mio Matsueda (Uni Oxford)

Correlations in monthly forecasts

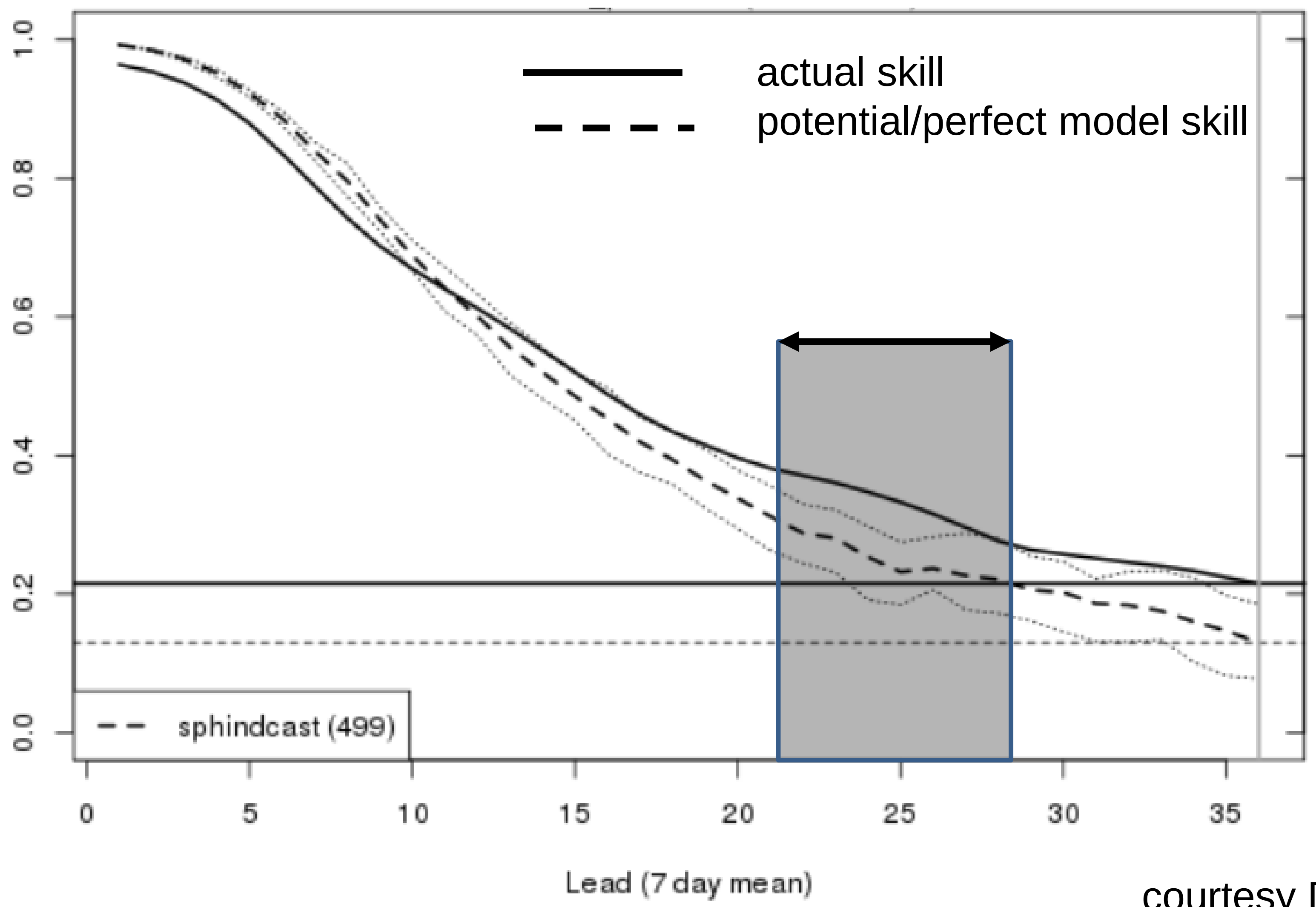
1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60 atmosphere only experiments with observed SSTs



courtesy Dan Rowlands (Cumulus)

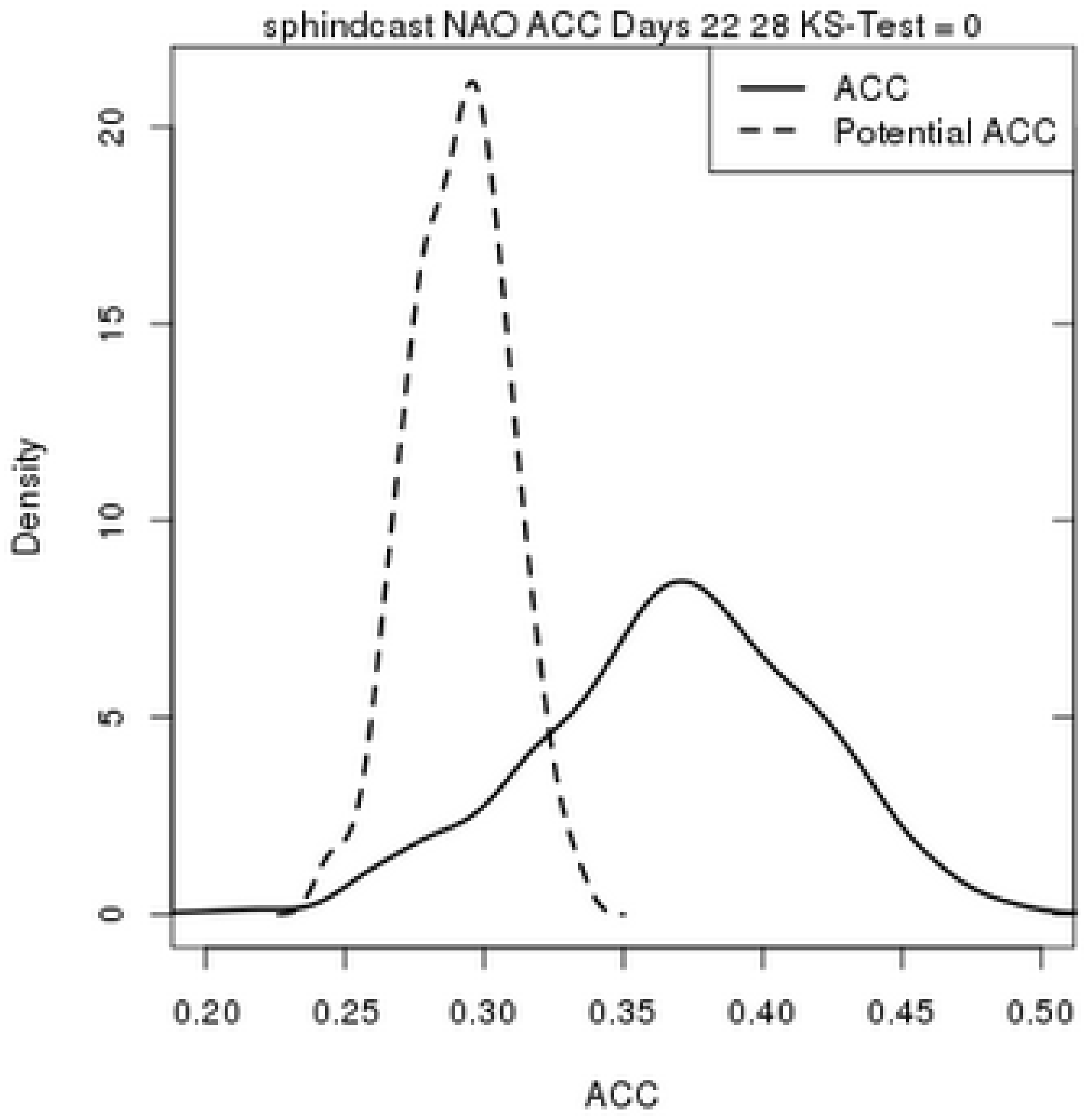
Correlations in monthly forecasts

1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60 atmosphere only experiments with observed SSTs



courtesy Dan Rowlands (Cumulus)

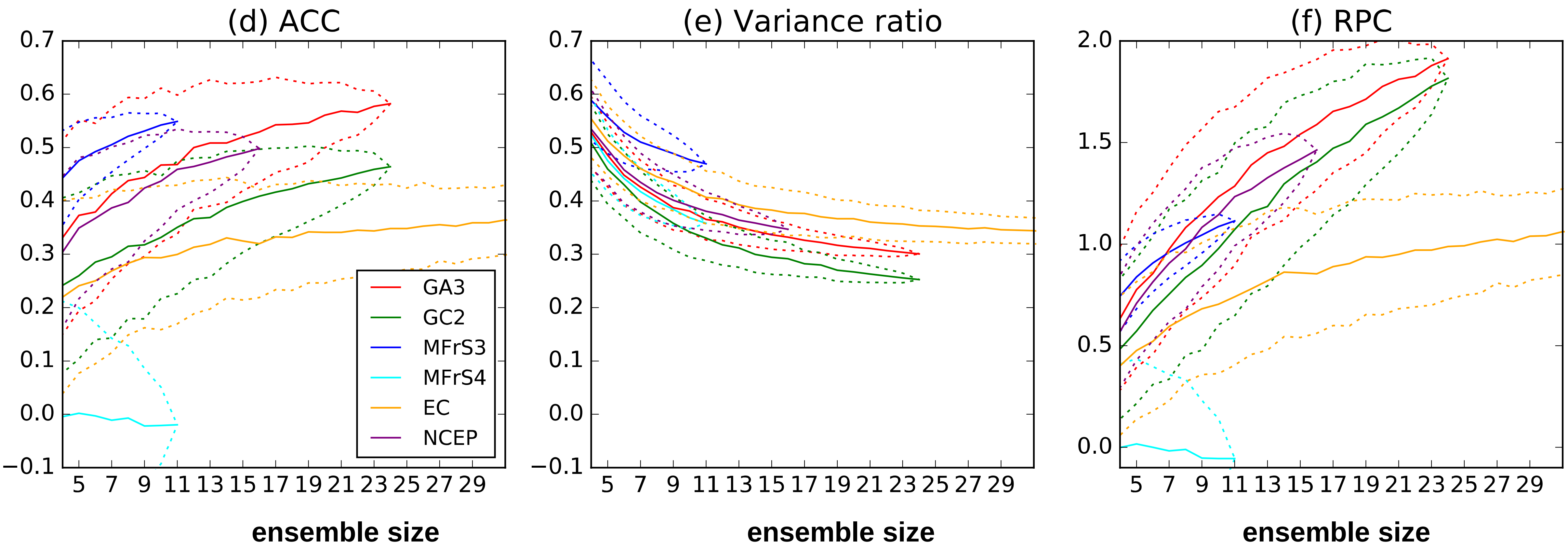
Correlations in week 4



courtesy Dan Rowlands (Cumulus)

Seasonal NAO predictions in the EUROSIP models

$$RPC = \frac{r}{\sqrt{VAR_{signal}/VAR_{total}}}$$



Met Office GloSea5-GA3 Met Office GloSea5-GA6 Météo France S3 Météo France S4 ECMWF S4 NCEP S2

courtesy Laura Baker (Uni Reading)

Perfect model definition

Should the perfect model ensemble mean include or exclude the verifying ensemble member?

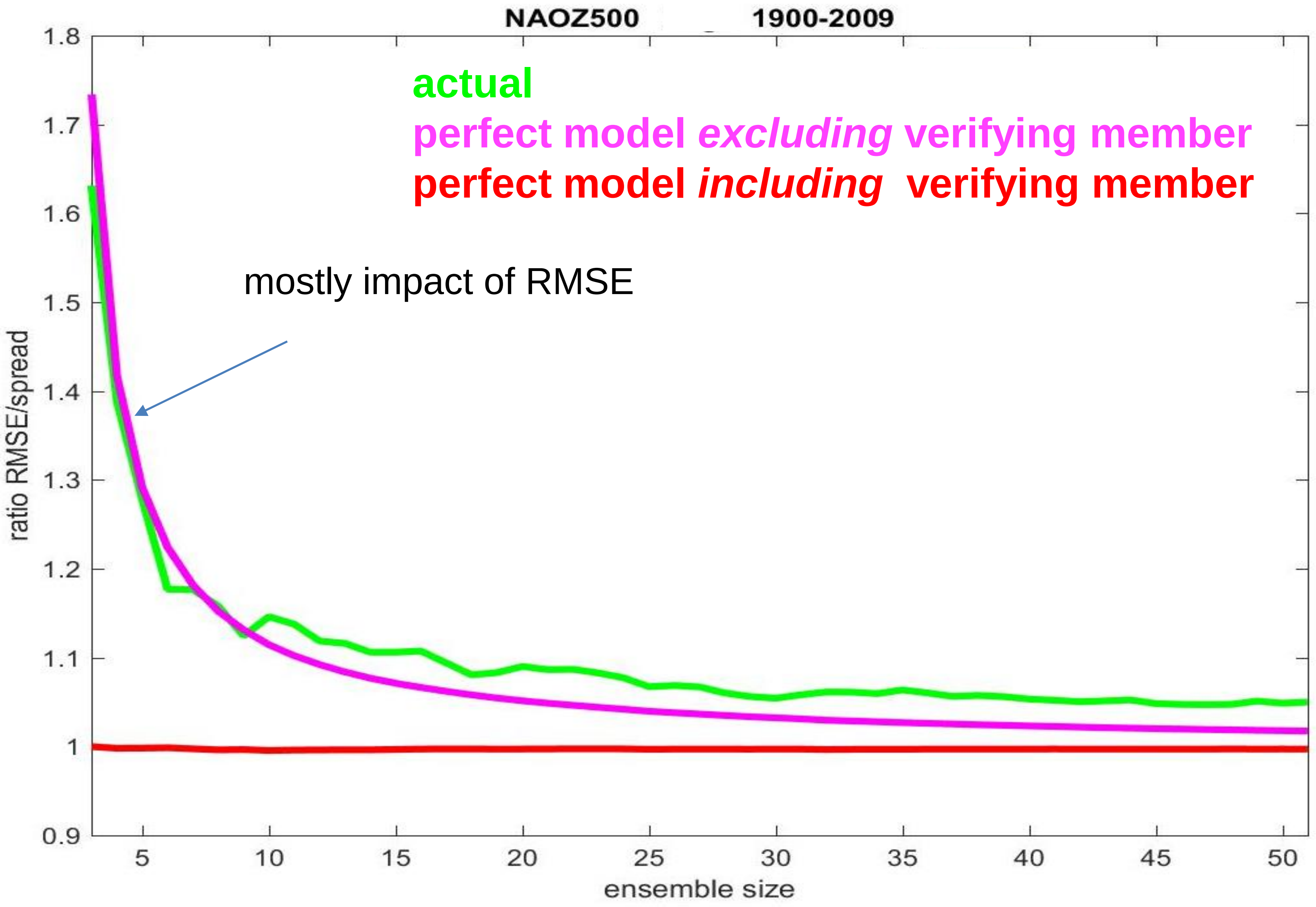
Most people would say “exclude” (forecasting context)

but

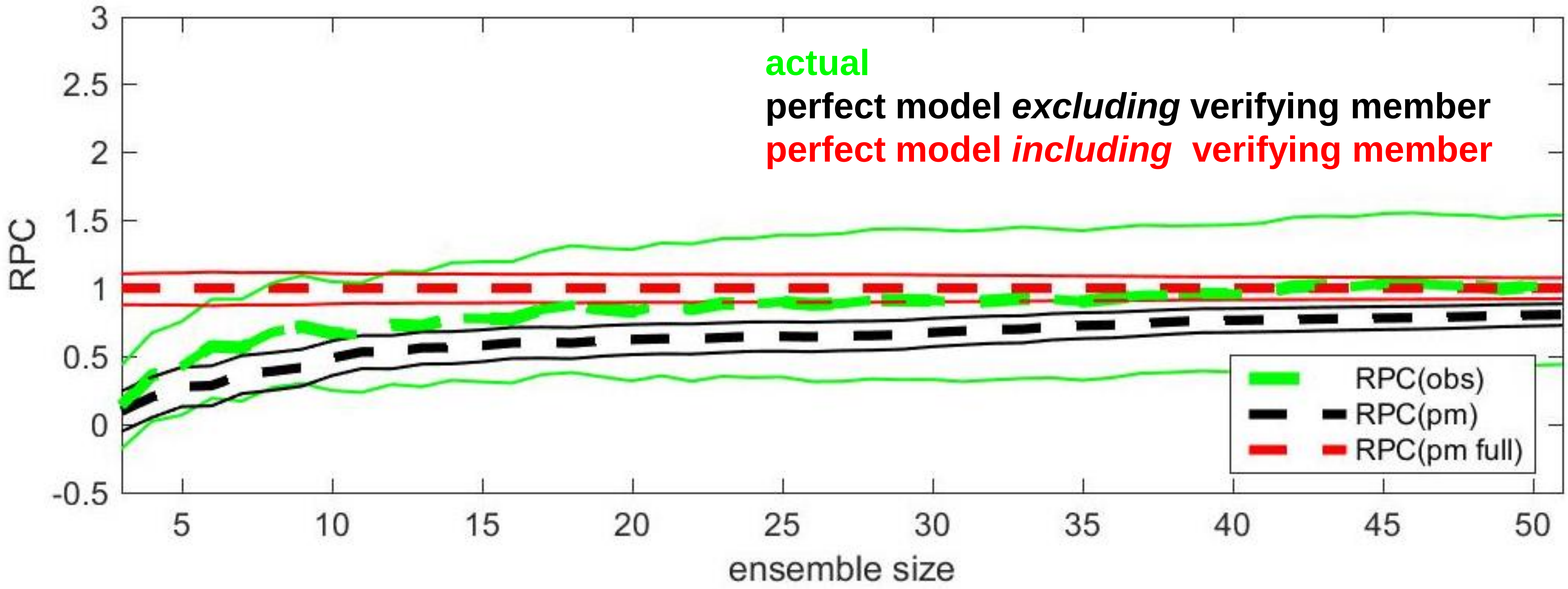
spread == RMSE and *RPC ==1*

are only fulfilled if the verifying member is included in the estimation of the ensemble mean.

Perfect model definition: Impact of the verifying ensemble member
Spread-RMSE relationship

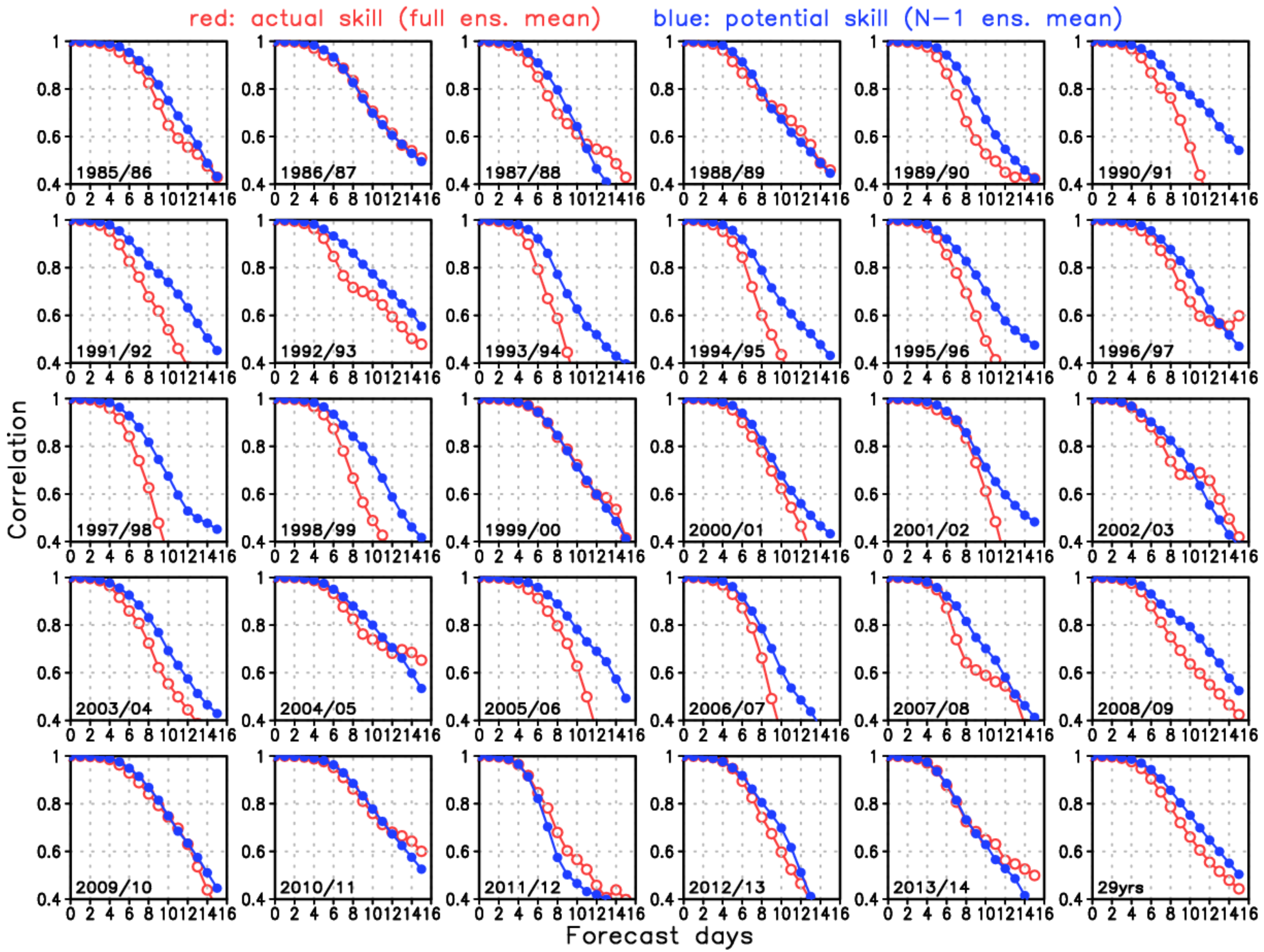


Perfect model definition: Impact of the verifying ensemble member
RPC



Perfect model definition: Impact of the verifying ensemble member

GEFS (NCEP) 30-year hindcasts



real world

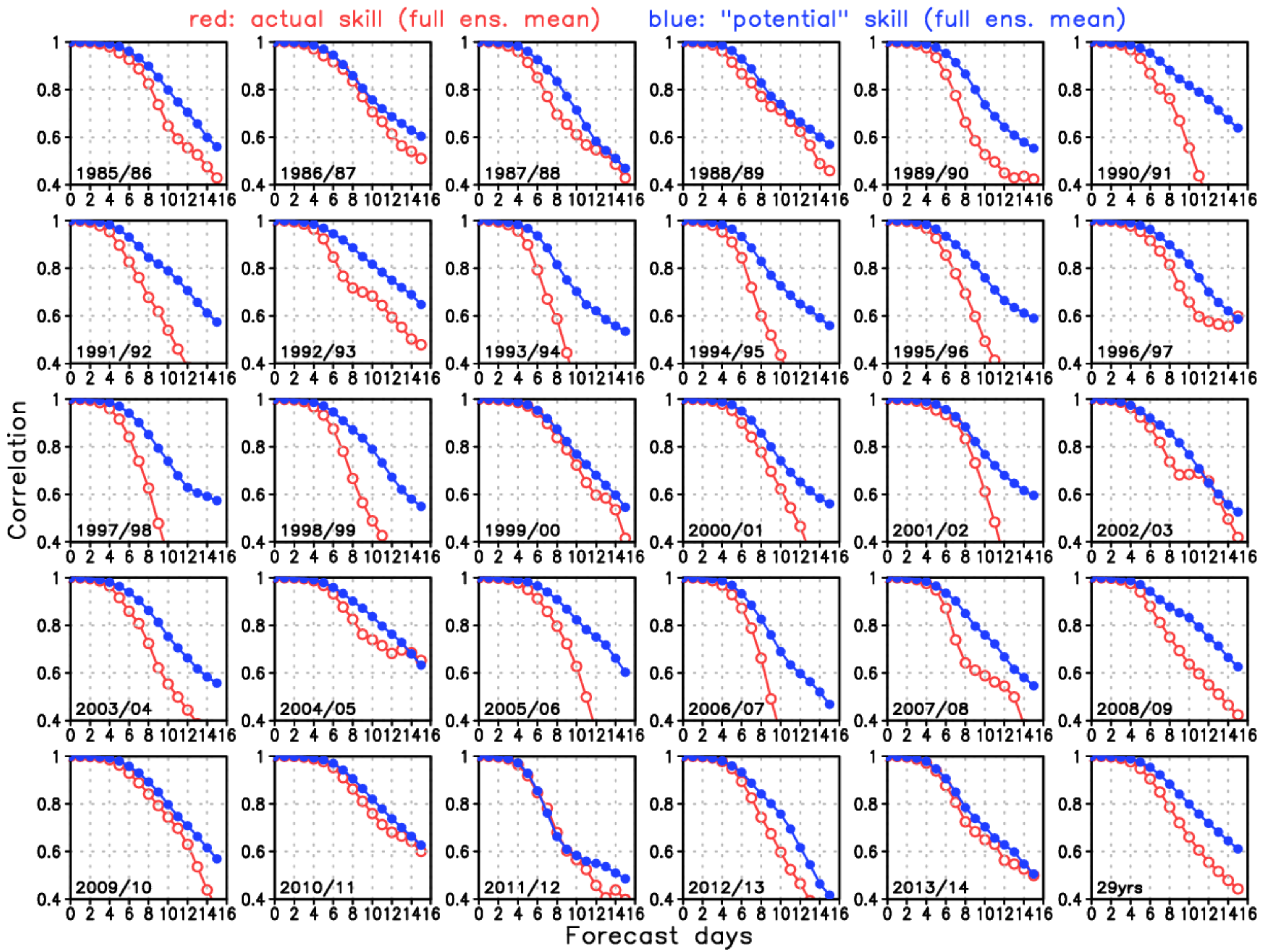
model world (potential or perfect model skill)

excluding verifying member

courtesy Mio Matsueda (Uni Oxford)

Perfect model definition: Impact of the verifying ensemble member

GEFS (NCEP) 30-year hindcasts



real world

model world (potential or perfect model skill)

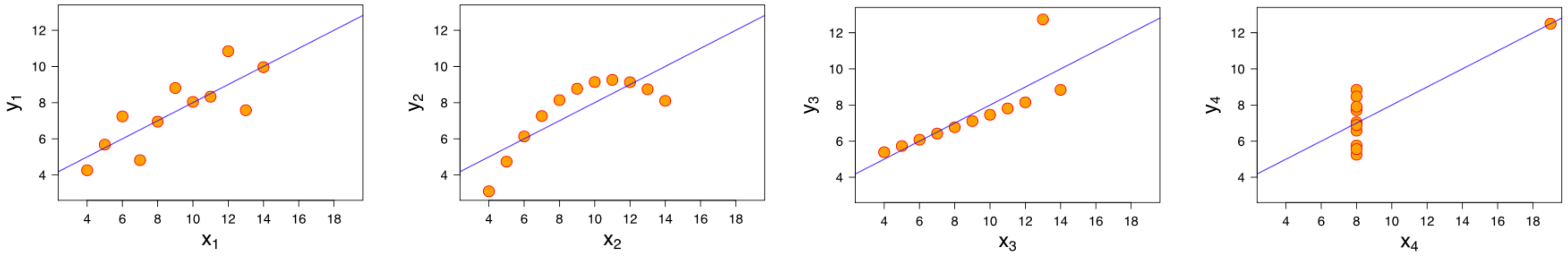
including verifying member

courtesy Mio Matsueda (Uni Oxford)

Illustrative example of correlation drawbacks after Anscombe (1973):

- Four pairs of x-y variables
- The four y variables have the same mean (=7.5), variance (=4.1) and **correlation (=0.82)**
- However, distributions of variables are very different

Anscombe's quartet



Normally distributed, "well behaved"

Not normally distributed, non-linear relationship

Perfect linear relationship, Except for one outlier

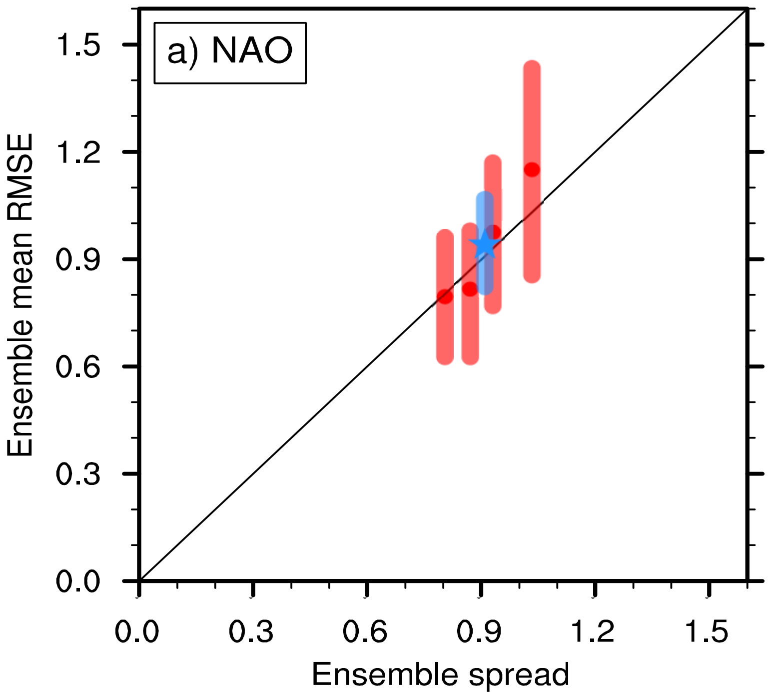
No relationship with one outlier

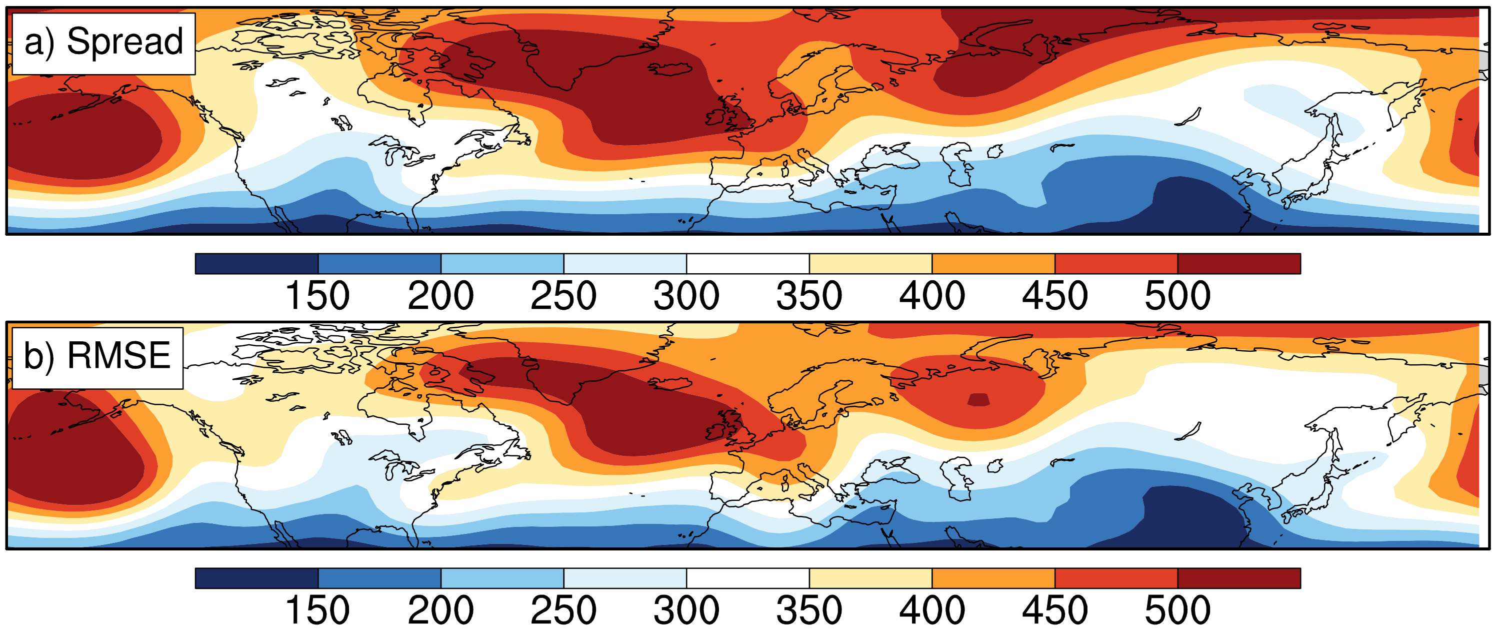
Anscombe (Amer. Statist. 1973)

Atmospheric Seasonal Forecasts of the 20th Century (ASF-20C)

- A new very long data set of seasonal hindcasts to study changes in predictability
- Use of ECMWF's re-analysis of the 20th Century (ERA-20C) that spans the 110-year period 1900 to 2010 to initialise atmospheric seasonal forecasts with ECMWF's forecast model
- SSTs and sea-ice are prescribed using HadISSTs
- Seasonal re-forecast experiments over the period 1900-2010
- Large ensemble of 51 perturbed members
- Focus here: 4-month forecast initialised on 1st of Nov each year to cover boreal winter (DJF) season
- More details in *Weisheimer et al. (QJRMS 2017)* and *O'Reilly et al. (GRL 2017)*

Spread-RMSE relationship in ASF-20C

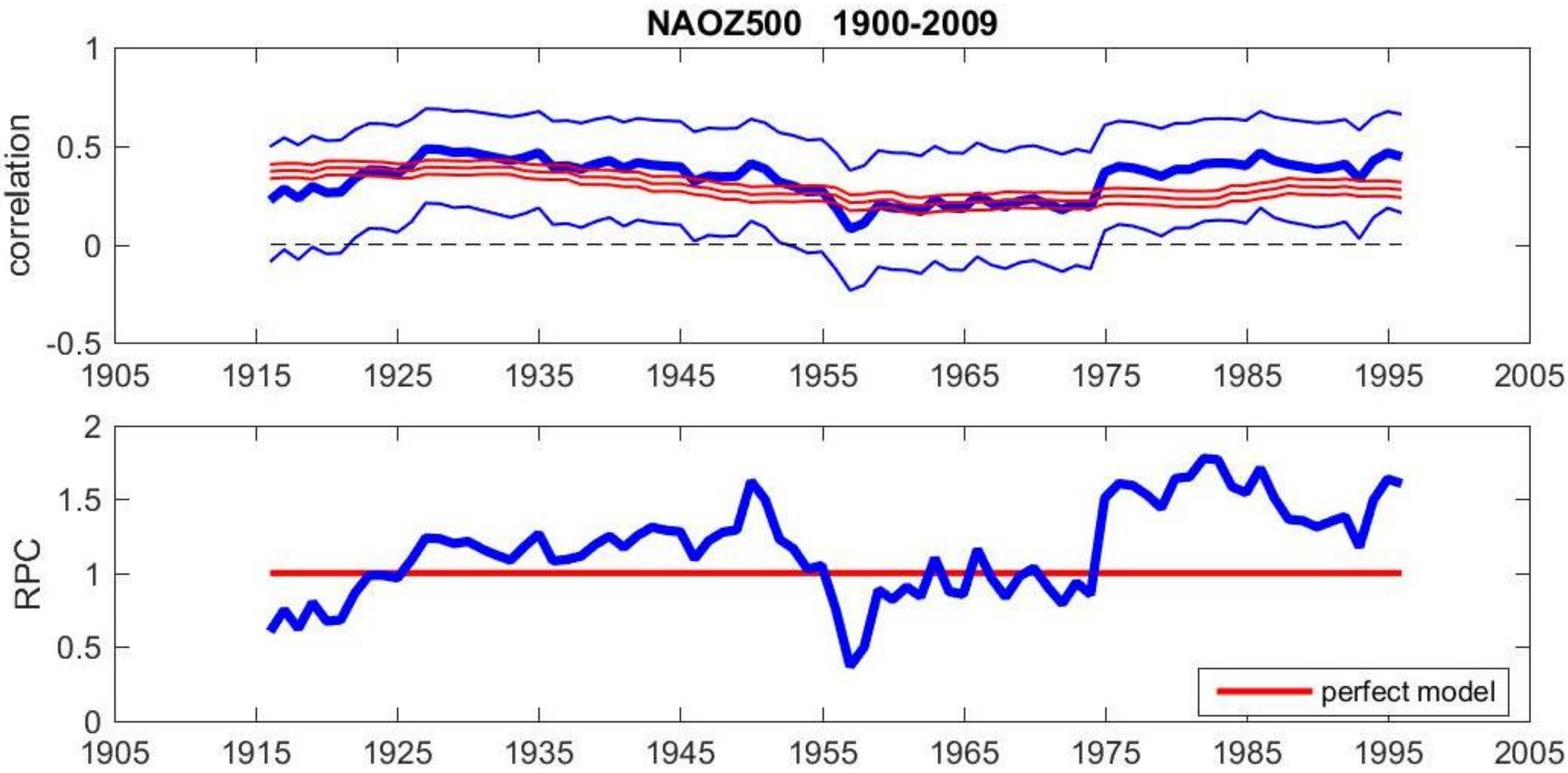




courtesy Dave MacLeod (Uni Oxford)

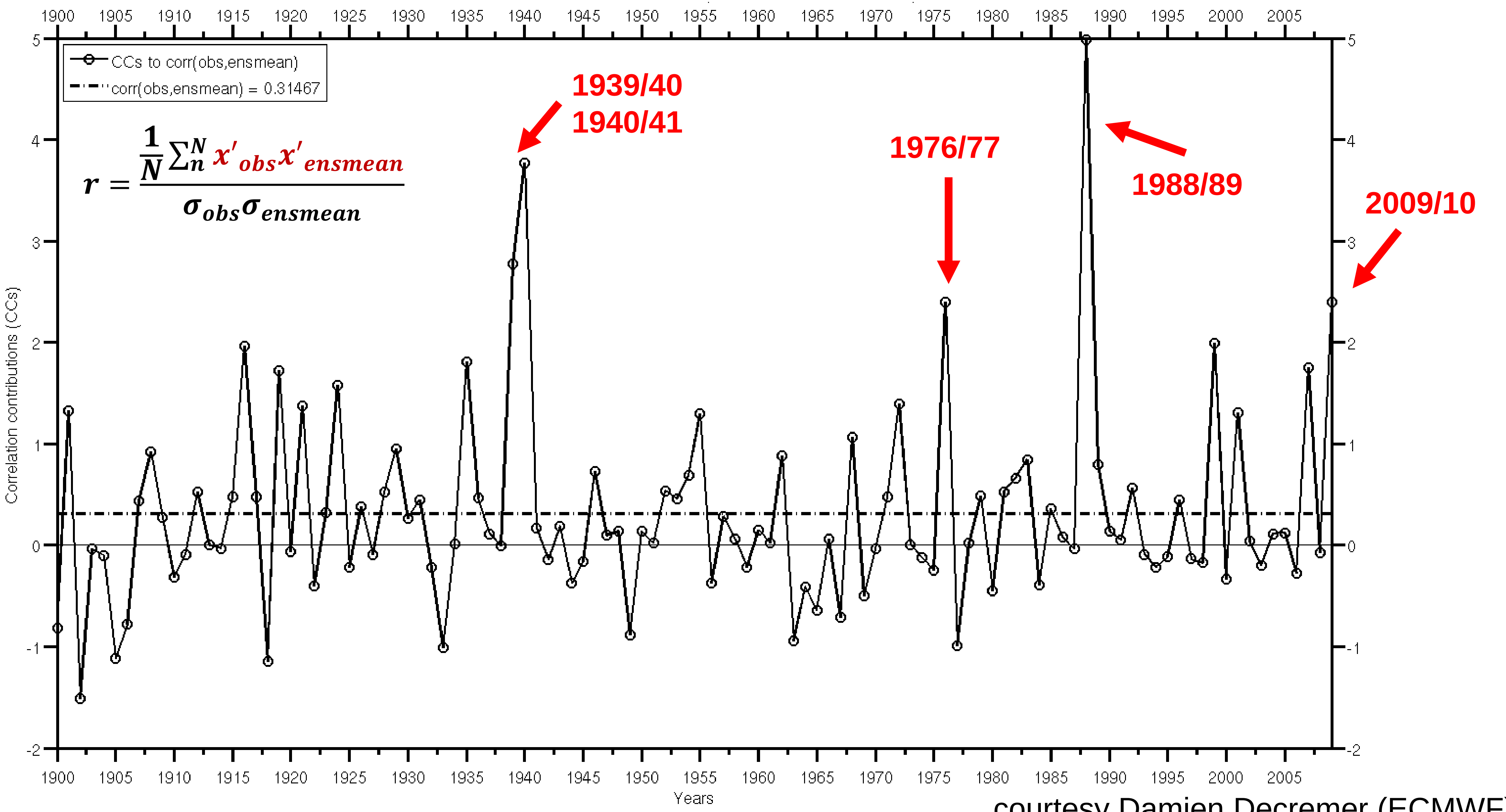
NAO skill and RPC in ASF-20C

using 30-year moving windows across the 110-year period



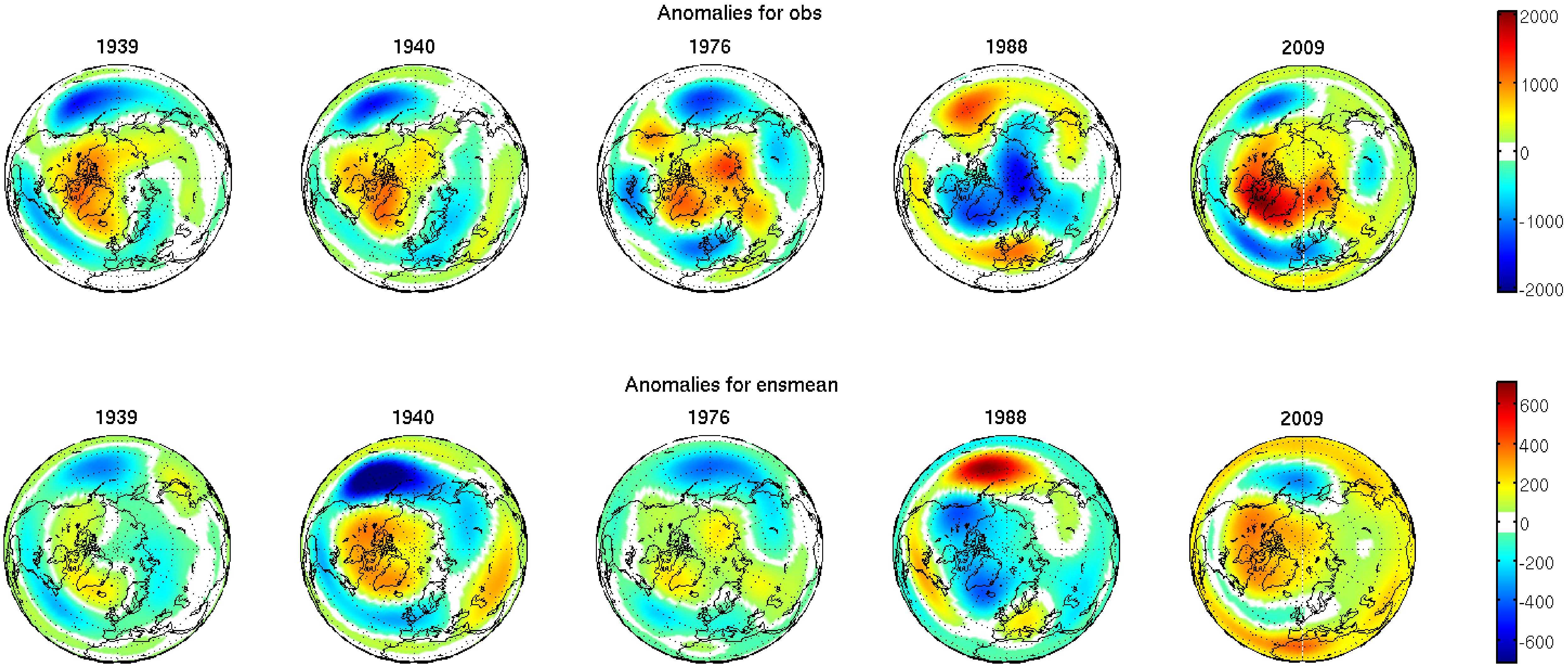
Weisheimer et al. (QJRMS 2017)

Contributions to covariance in ASF-20C



courtesy Damien Decremer (ECMWF)

Z500 anomalies for largest contributions to covariance

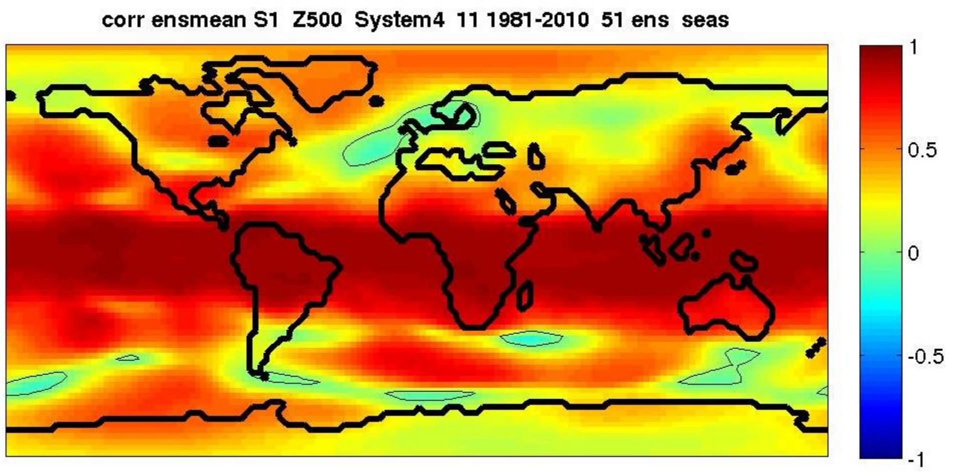


courtesy Damien Decremer (ECMWF)

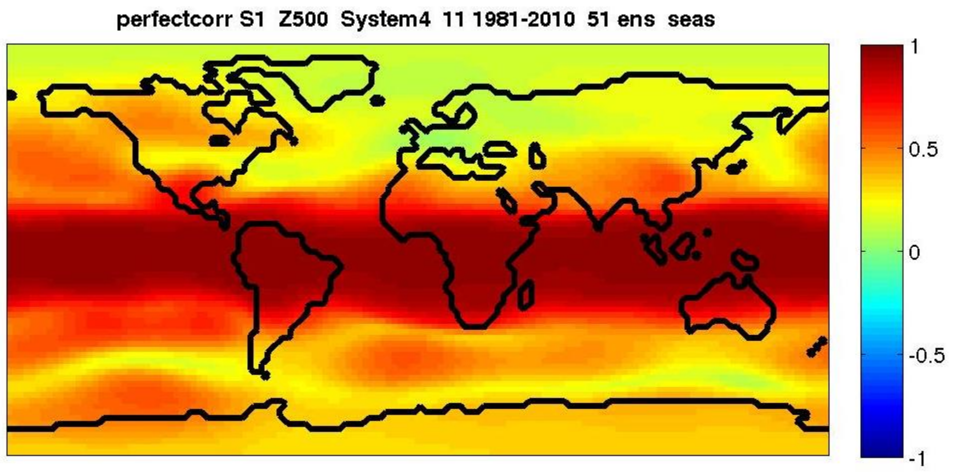
1st Nov start date 1981-2010 Z500 seasonal forecasts S4 51 ens members

DJF mean

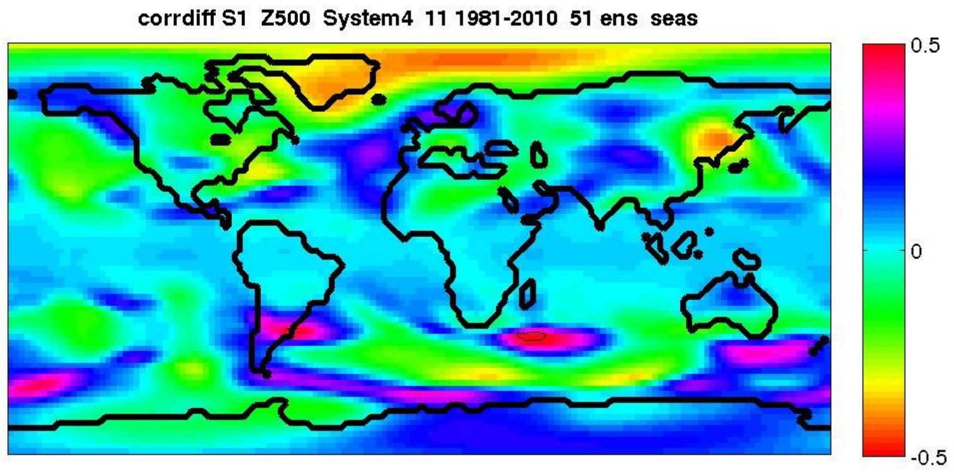
correlation skill



perfect model correlation skill



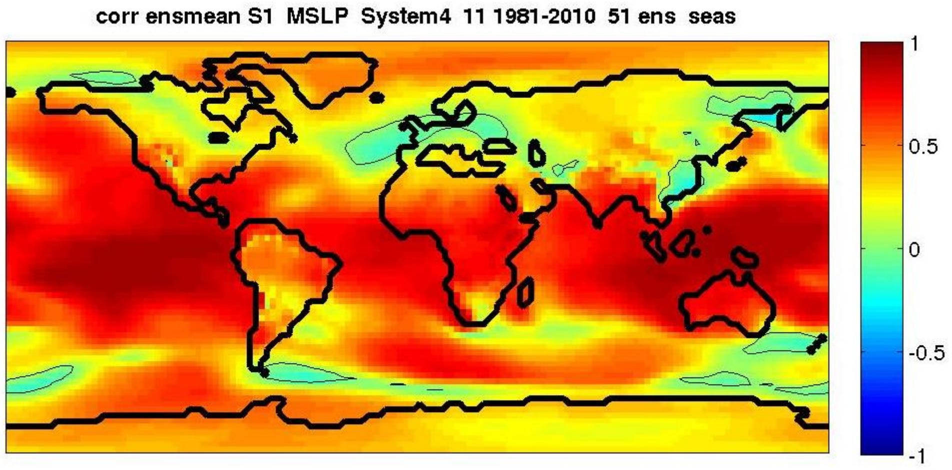
correlation skill *minus* perfect model correlation skill



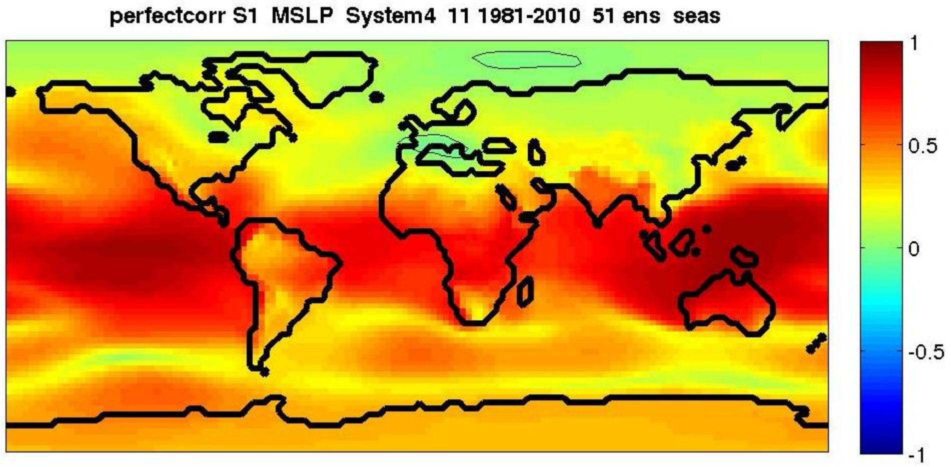
1st Nov start date 1981-2010 MSLP seasonal forecasts S4 51 ens members

DJF mean

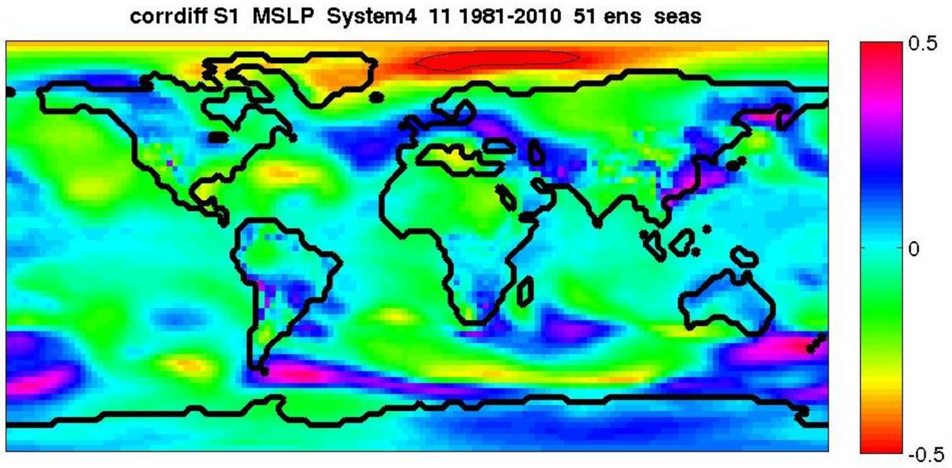
correlation skill



perfect model correlation skill



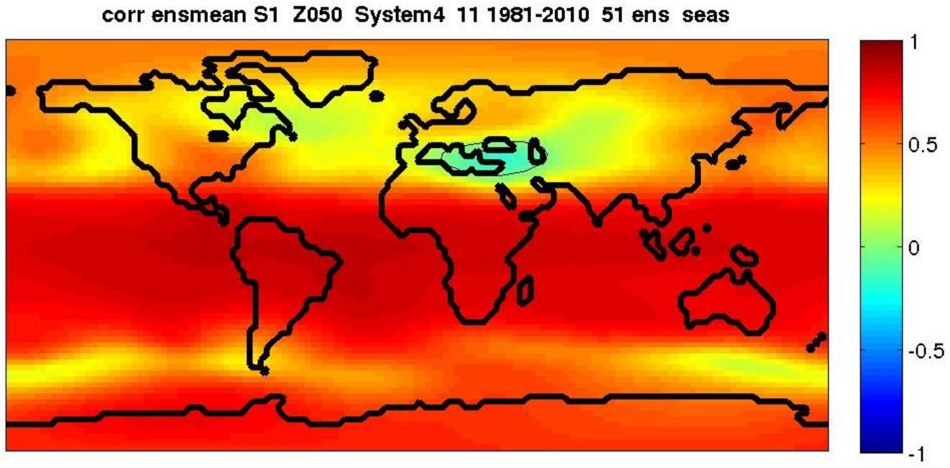
correlation skill *minus* perfect model correlation skill



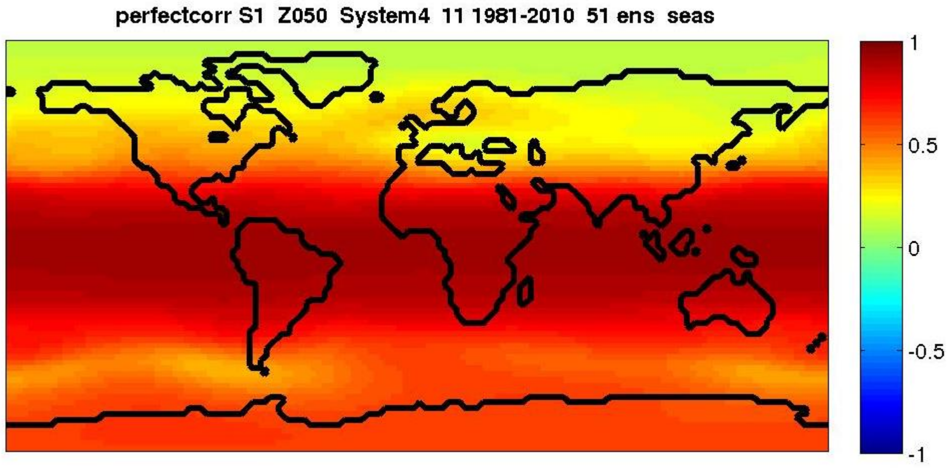
1st Nov start date 1981-2010 Z50 seasonal forecasts S4 51 ens members

DJF mean

correlation skill



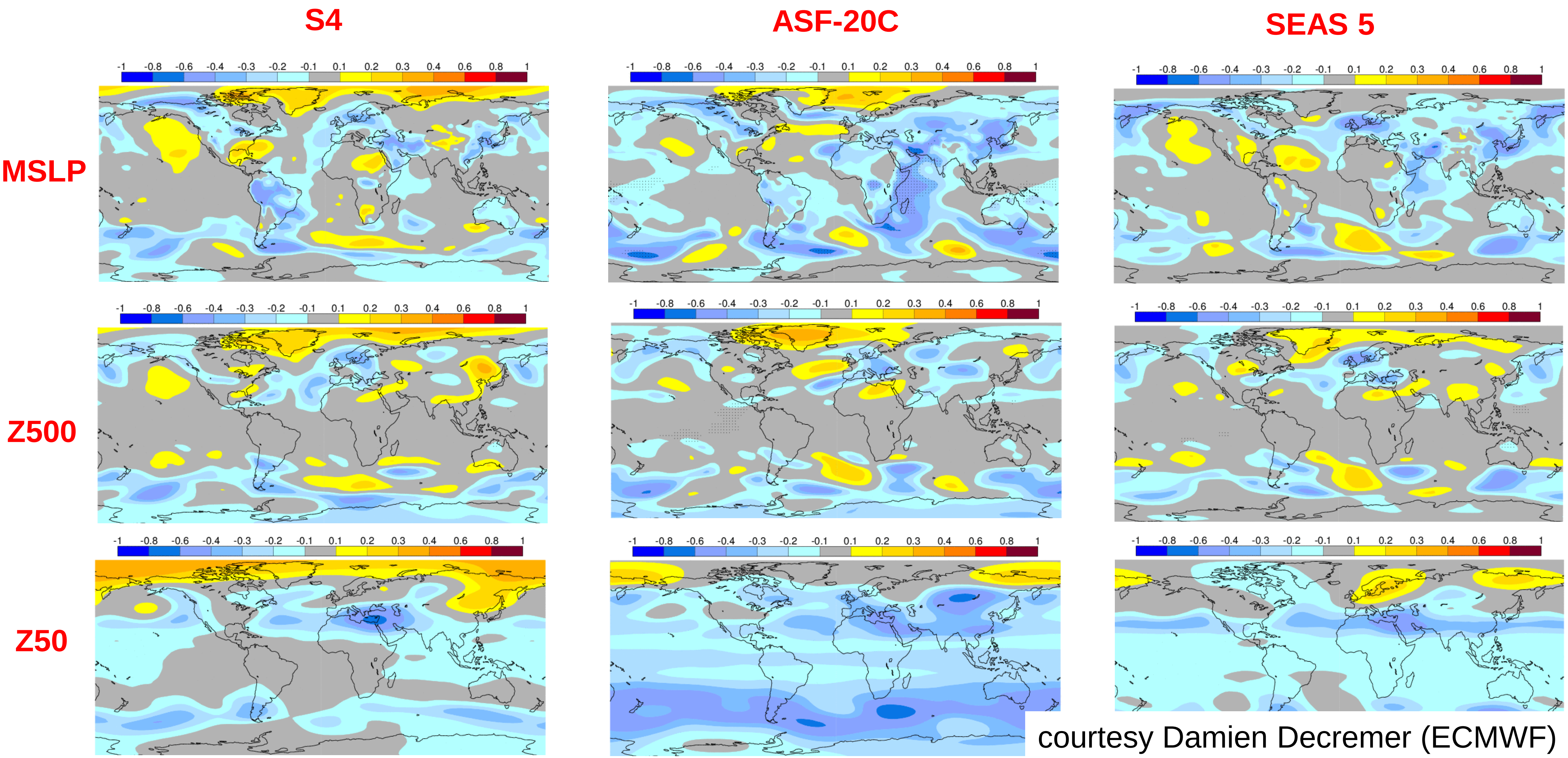
perfect model correlation skill



1981-2009

Underconfidence in seasonal forecasts?

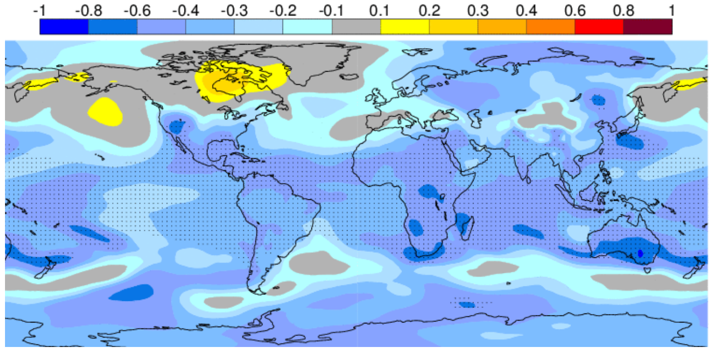
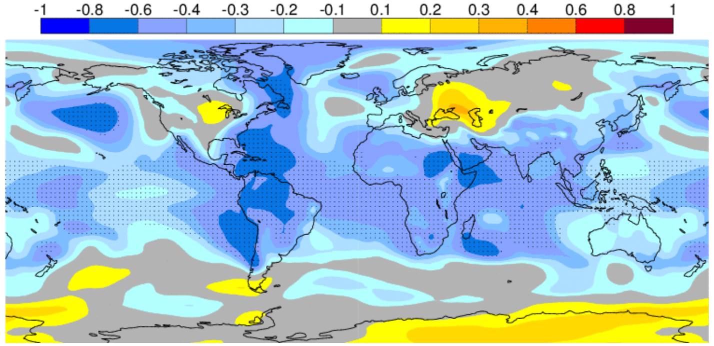
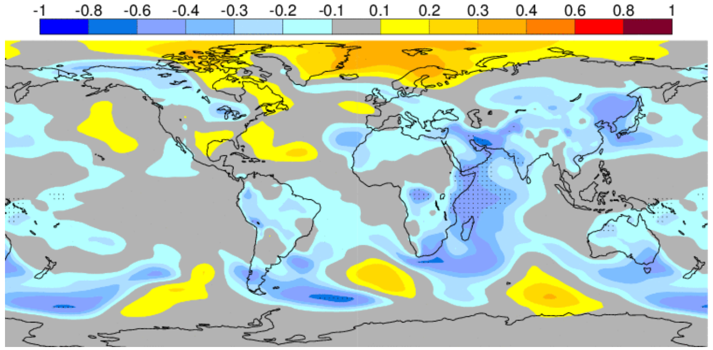
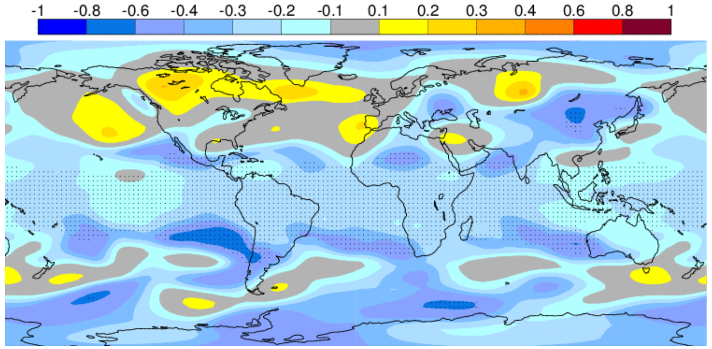
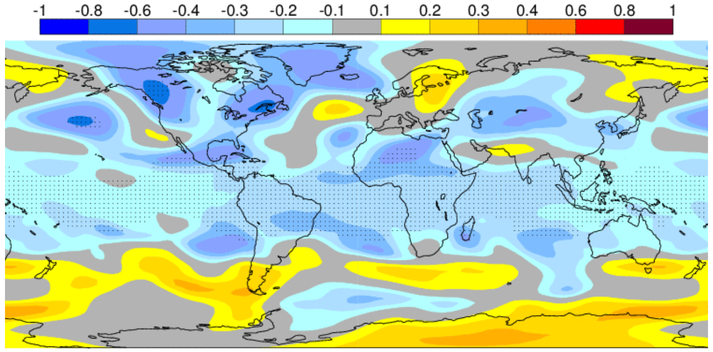
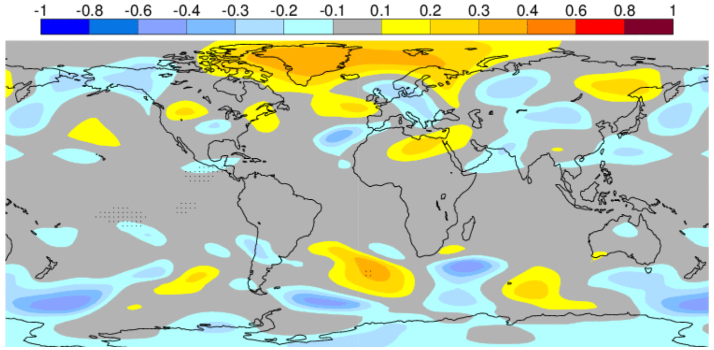
corr(obs,ensmean) minus *corr(ens,ensmean)*



courtesy Damien Decremer (ECMWF)

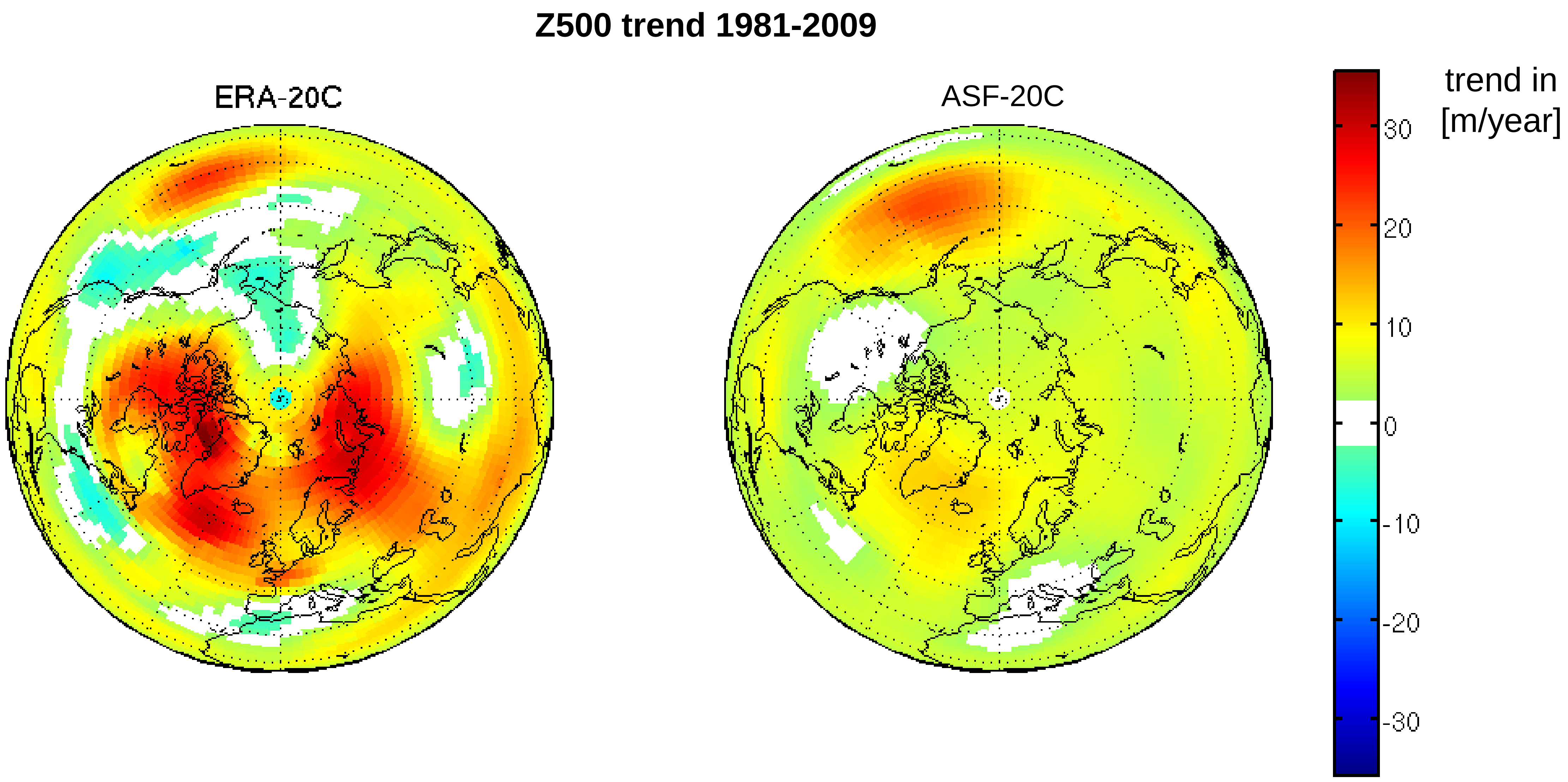
Underconfidence in seasonal forecasts (ASF-20C)?

corr(obs,ensmean) minus *corr(ens,ensmean)*

	1912-1940	1942-1970	1981-2009
MSLP			
Z500			

courtesy Damien Decremer (ECMWF)

Arctic amplification?



courtesy Damien Decremer (ECMWF)

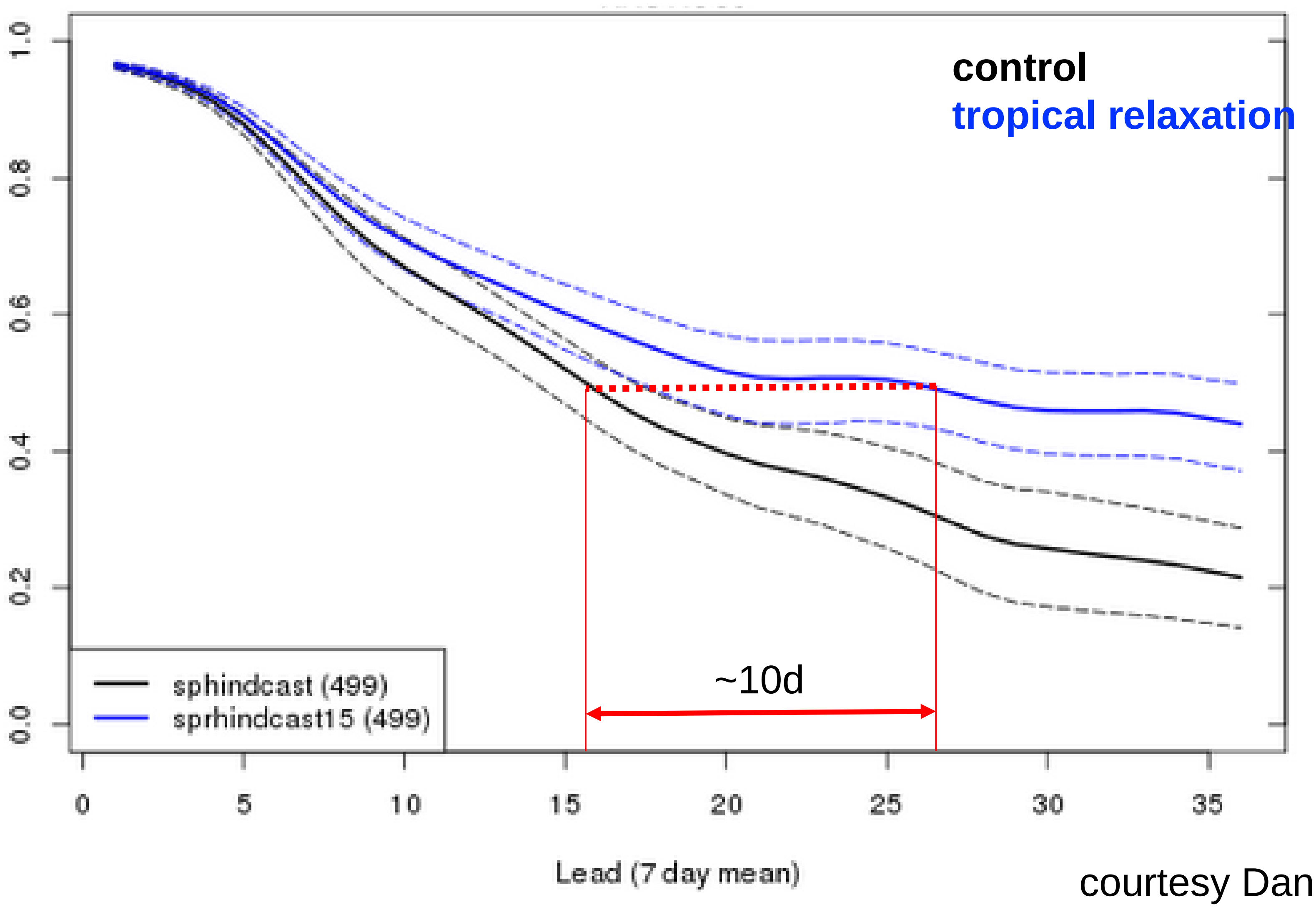
Role of the Tropics

as a major source of predictability on longer timescale?

Anomaly Correlation of the NAO in monthly forecast experiments

1995/96 – 2016/17 hindcasts with 11 ensemble members

CY41R1 T255L60 atmosphere only experiments with observed SSTs

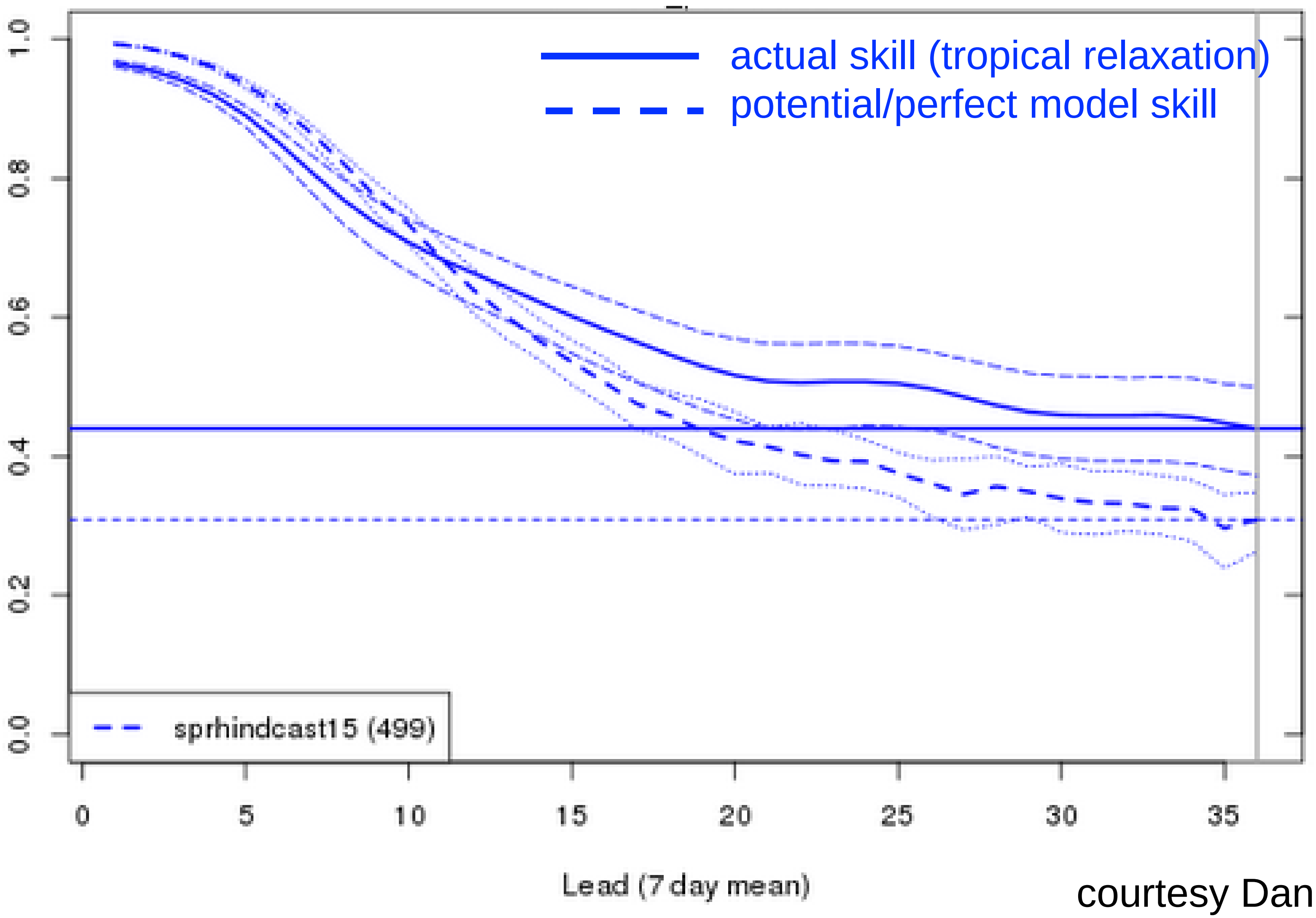


courtesy Dan Rowlands (Cumulus)

Anomaly Correlation of the NAO in monthly forecast experiments

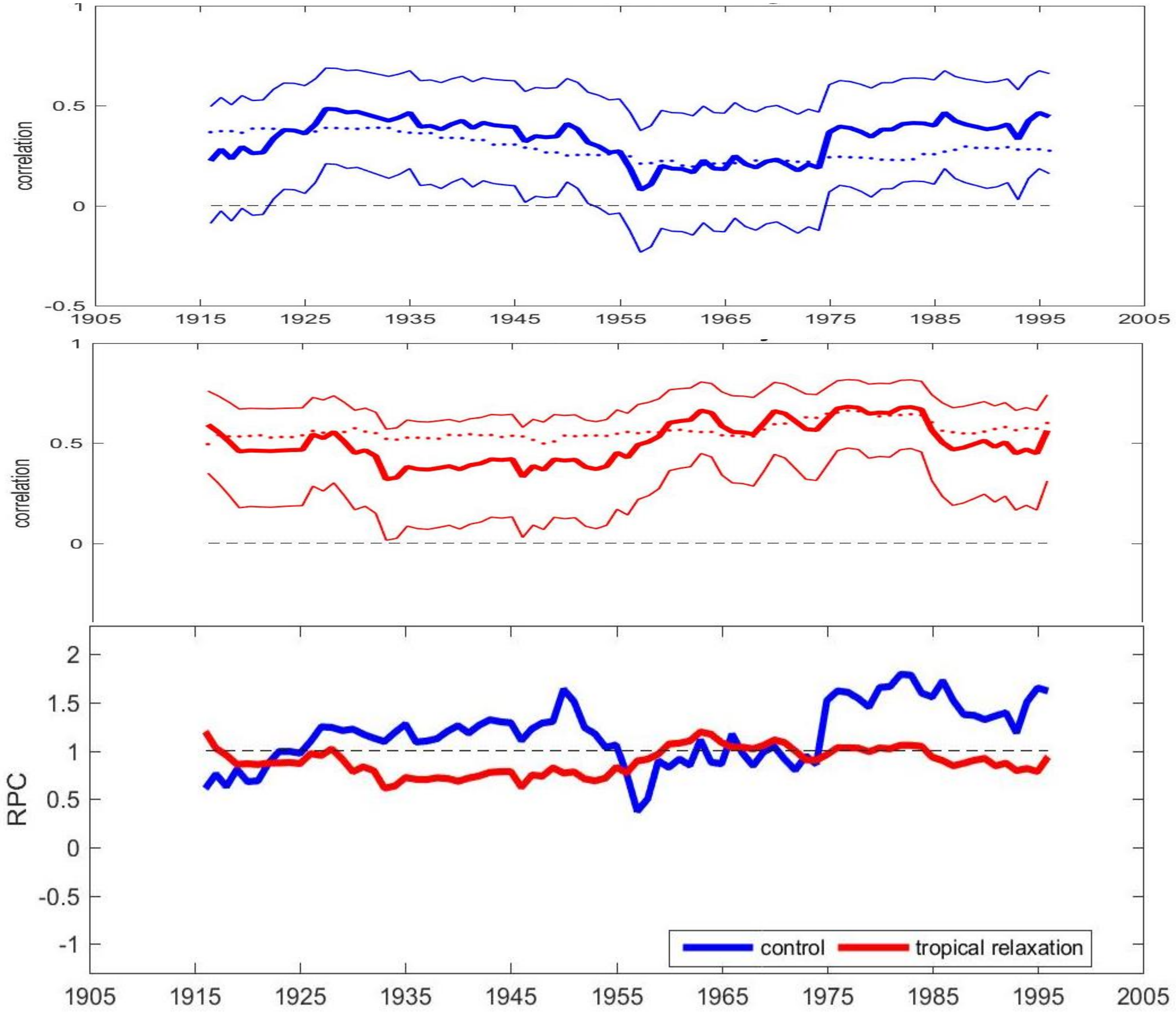
1995/96 – 2016/17 hindcasts with 11 ensemble members

CY41R1 T255L60 atmosphere only experiments with observed SSTs



courtesy Dan Rowlands (Cumulus)

Role of the Tropics in ASF-20C



corr(obs,ensmean)

corr(ens,ensmean) = perfect model

control

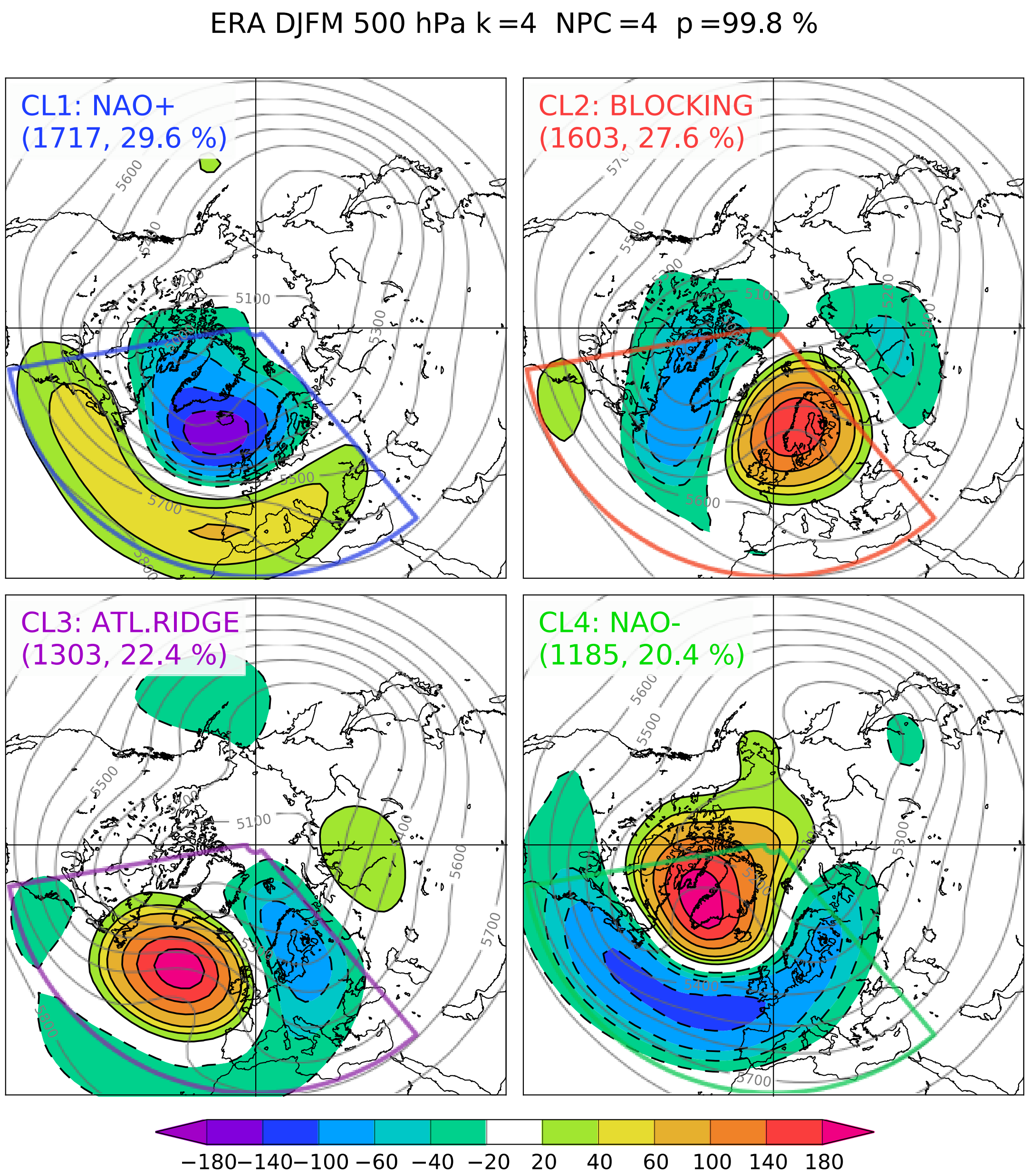
tropical relaxation

Why is RPC > 1 during decadal periods when NAO > 0 ?

→ Tim Palmer's regime hypothesis

Circulation regimes over the Euro-Atlantic area

High Zonal Index



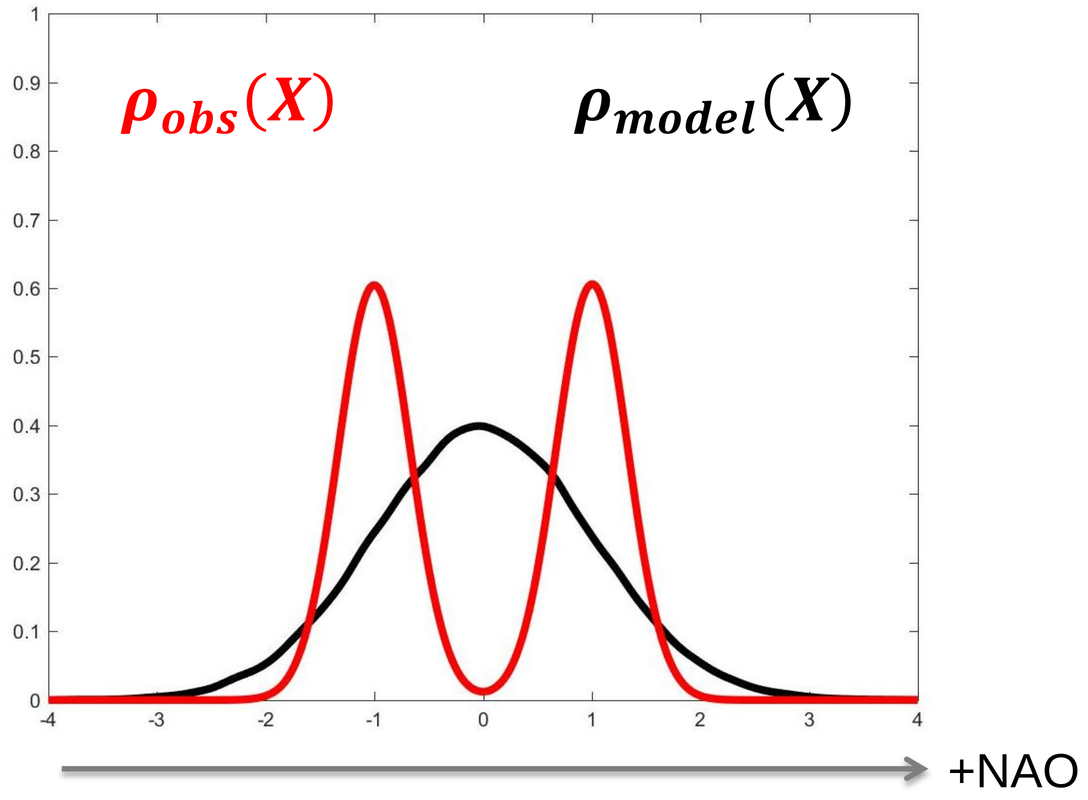
Low Zonal Index

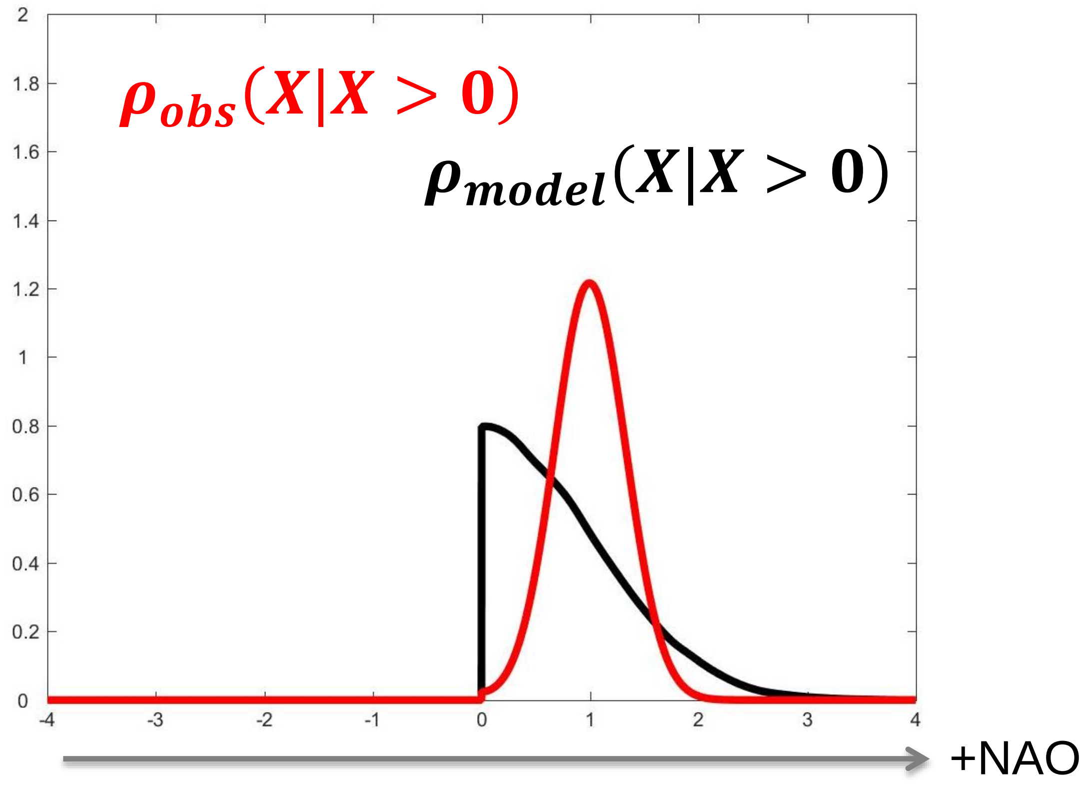
Low Zonal Index

Low Zonal Index

Dawson et al. (GRL 2012)

Effect of non-linear regime error





$$\mu_{model} = \mu_{obs} = 0$$

$$\sigma_{model} = \sigma_{obs} = 1$$

Model's first two moments agree perfectly with observations

$$\mu_{obs} = 1.0 \quad \mu_{model} = 0.8$$

$$\sigma_{obs} = 0.3 \quad \sigma_{model} = 0.6$$

$$\sigma_{model} > \sigma_{obs}$$

courtesy Tim Palmer (Uni Oxford)

Summary and Conclusions

It has recently been suggested that **predictability estimates of seasonal forecast models of the winter NAO underestimate the real world predictability**. These findings are based on multi-decadal simulations when the NAO was predominantly in its positive phase.

Spread-RMSE diagnostics across forecast time scales give **no indication of over-dispersive behaviour**. **Correlation skill** does indicate situations with **perfect model skill > actual skill** on time scales of ~14d onwards.

However, correlation measures suffer from **large uncertainties due to small samples** taken over specific long-term (decadal-centennial) climate regimes.

Long seasonal hindcasts covering the full 20th Century have recently become available and allow to put the predictability situation of the recent decades into a longer climate context. Over the entire period **RPC~1**.

Recent decades see **high levels of NAO skill and a tendency to underestimate the real skill**. Previous climate periods do not show indications for such a behaviour.

Preferred flow pattern of most skillful years point towards **strong Z500 anomalies over Greenland and parts of the Artic.** Observed **predictability is higher throughout the atmospheric column** in these regions but **only during recent decades**.

Role of Tropics as a major source of seasonal predictability is yet controversial. → link to teleconnections?

“Conundrum” (or paradox) is a plausible manifestation of **model deficiencies in representing non-linear circulation regimes**.

References

Anscombe, F.J. (1973). Graphs in Statistical Analysis. *American Statistician*. **27** (1): 17–21.

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson (2014). Do seasonal-to- decadal climate predictions under- estimate the predictability of the real world? *Geophys. Res. Lett.,* **41**, 5620–5628, doi:10.1002/2014GL061146.

Kumar, A. (2009). Finite samples and uncertainty estimates for skill measures for seasonal prediction, *Mon. Weather Rev.*, **137**, 2622–2631, doi:10.1175/2009MWR2814.1.

MacLeod, D., C. O'Reilly, T.N. Palmer, A. Weisheimer (2017). Flow dependent ensemble forecast spread in seasonal forecasts of the Northern Extratropical winter. In preparation.

Müller, W., C. Appenzeller and C. Schär. Probabilistic seasonal prediction of the winter North Atlantic Oscillation and its impact on near surface temperature. *Clim. Dyn.* (2005), **24**, 213-226, doi:10.1007/s00382-004-0492-z.

Scaife, A. A., et al. (2014), Skilful long range prediction of European and North American winters, *Geophys. Res. Letts.,* **41**, 2514–1519, doi:10.1002/2014GL059637.

Shi, W., N. Schaller, D. MacLeod, T.N. Palmer and A. Weisheimer. Impact of hindcast length on estimates of seasonal climate predictability. *Geophys. Res. Lett.* (2015), **42**,doi:10.1002/2014GL062829.

Siegert, S., D.B. Stephenson, P.G Sansom, A.A. Scaife, R. Eade and A. Arribas. A Bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is the NAO predictability? *J. Climate* (2016), **29**, 995-1012.

Stockdale, T. N., F. Molteni, and L. Ferranti (2015), Atmospheric initial conditions and the predictability of the Arctic Oscillation, *Geophys. Res. Lett.,* **42**, 1173–1179, doi:10.1002/ 2014GL062681.

O'Reilly, C. H., J. Heatley, D. MacLeod, A. Weisheimer, T. N. Palmer, N. Schaller, and T. Woollings (2017), Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century, *Geophys. Res. Lett.,* **44**, 5729–5738, doi:10.1002/2017GL073736.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A. and Palmer, T. (2017), Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Q.J.R. Meteorol. Soc,* **143**: 917–926. doi:10.1002/qj.2976