Assessing the skill of SEAS5

Antje Weisheimer

Antje.Weisheimer@ecmwf.int

With thanks to Stephanie Johnson and Tim Stockdale

Climate
Change Service

climate.copernicus.eu

If you cannot predict the weather a week ahead, how can you predict the climate for the coming season?

Predictability of the first and second kind

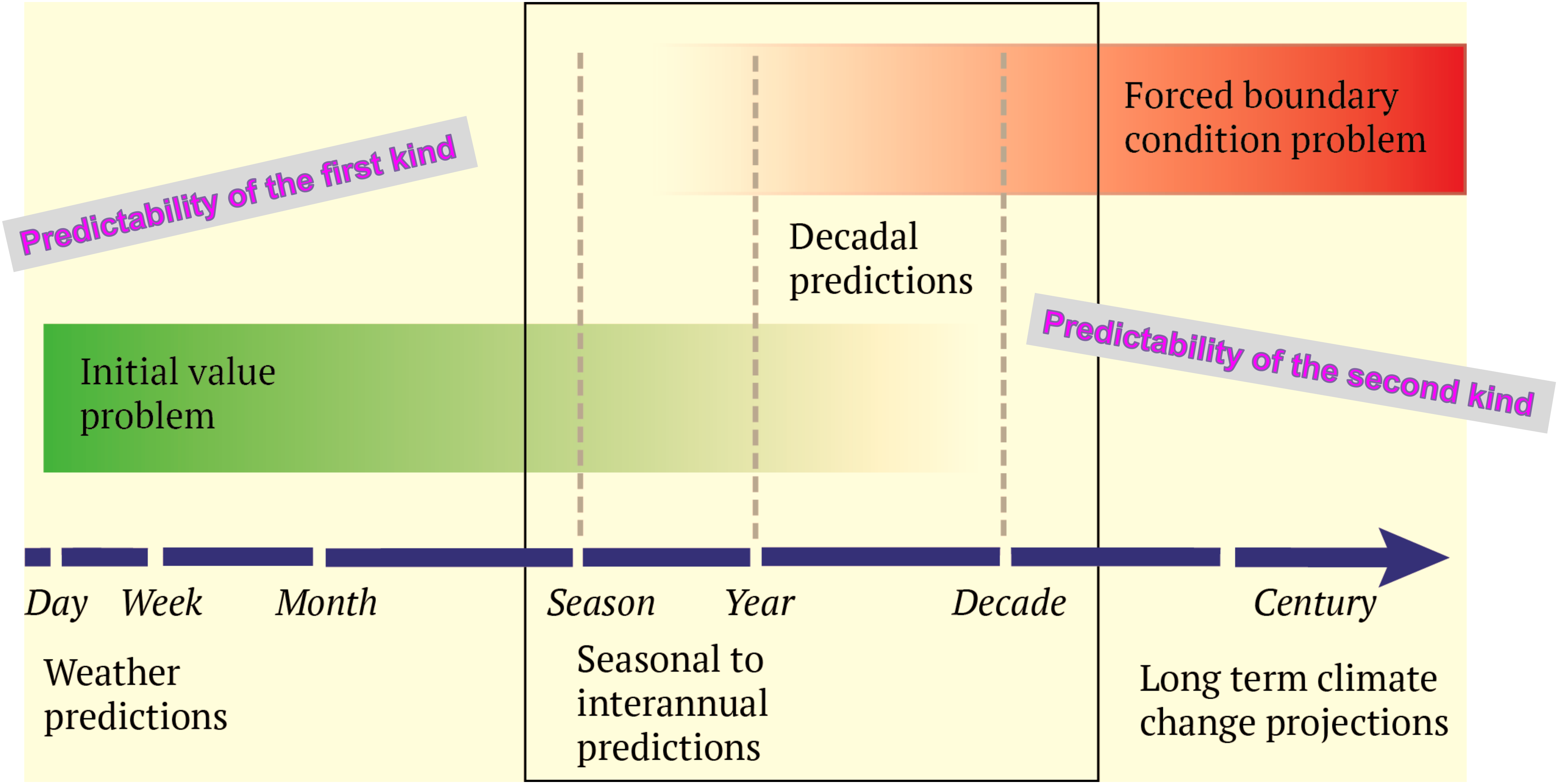
(Ed Lorenz: Climatic predictability, 1975)

- Numerical models used to predict weather and climate are based on the same physical principles
  e.g. Newton's laws of motion and the laws of thermodynamics

- However, forecasting the weather is a different problem to forecasting the climate

- **Weather prediction – *predictability of the first kind***
  - Predictability horizon is severely limited by the chaotic nature of the atmosphere → detailed evolution of the weather becomes unpredictable after O(10 days)
  - Predictability relies on sensitive dependence on initial conditions

- **Climate prediction – *predictability of the second kind***
  - Does not aim to forecast *individual weather* patterns
  - Rather, to forecast the *statistics* of the climate system averaged over time and space
  - Predictability relies on the sensitivity of the system to its boundary conditions (e.g. radiative forcing) and on the long memory of initial conditions of slowly evolving components of the climate system (e.g. the ocean, land surface)
  - Predictability horizon for climate forecasts is much longer than for weather forecasts


Predictability of the first kind
Forced boundary condition problem
Decadal predictions
Predictability of the second kind
Initial value problem
Day
Week
Month
Season
Year
Decade
Century
Weather predictions
Seasonal to interannual predictions
Long term climate change projections

Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting

J. Shukla

The Earth's atmosphere is generally considered to be an example of a chaotic system that is sensitively dependent on initial conditions. It is shown here that certain regions of the atmosphere are an exception. Wind patterns and rainfall in certain regions of the tropics are so strongly determined by the temperature of the underlying sea surface that they do not show sensitive dependence on the initial conditions of the atmosphere. Therefore, it should be possible to predict the large-scale tropical circulation and rainfall for as long as the ocean temperature can be predicted. If changes in tropical Pacific sea-surface temperature are quite large, even the extratropical circulation over some regions, especially over the Pacific–North American sector, is predictable.

At the beginning of the 20th century it was hypothesized that it should be possible to predict weather by solving the mathematical equations that describe the physical laws that govern the motion of air. It took several decades to develop an appropriate set of

that aspects of the tropical atmosphere do not conform to the above definition of chaos. The tropical flow patterns and rainfall, especially over the open ocean, are so strongly determined by the underlying sea-surface temperature (SST) that they show little sensitivity to

predict large-scale changes in the winter season mean circulation over North America several months in advance, as indeed was the case for the 1997–1998 El Niño. However, the extent to which this apparent high potential predictability of the tropical and extratropical atmosphere can be realized in routine forecasting will depend on our ability to predict the SST itself.

The numerical model used in this research has been described (*3*). The dynamic equations and the numerical techniques used to integrate the model are the same as those used by the U.S. National Weather Service for routine weather prediction, and the accuracy of short-range weather forecasts made with this model is comparable to the state-of-the-art weather forecast models.

Two sets of simulations were carried out with the same prescribed SST but quite large differences in the initial conditions of the atmosphere. This simulation requires a selection of two very different initial conditions. Rather than choosing them arbitrarily, or constructing them artificially, atmospheric states observed during the past 50 years were chosen. The data show that the Southern Oscil-

Science (1998)

Aspects of the tropical atmosphere *do not* follow the deterministic chaos paradigm:

- Tropical flow and rainfall patterns, especially over the ocean, are strongly coupled to the underlying SSTs which show little sensitivity to initial conditions of the atmosphere
- Even with very large changes in the atmospheric IC (as large as the climatological variability), the resulting large-scale wind and rainfall pattern in certain tropical regions do not diverge

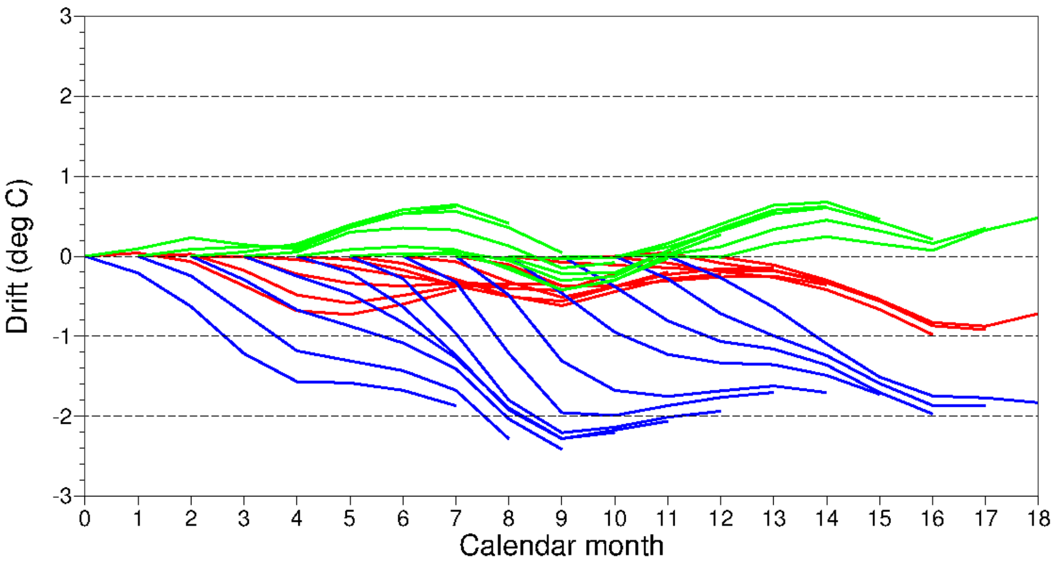
A unique and fundamental property of the tropical atmosphere:

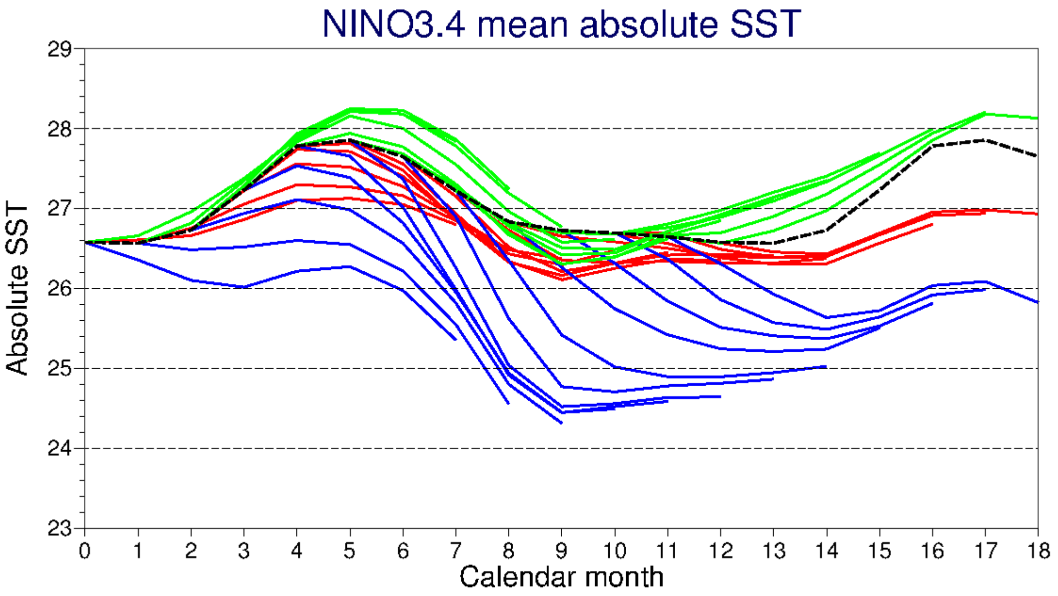
Tropical large-scale seasonal circulation and rainfall are almost completely determined by the boundary conditions of SSTs.

1. It should be possible to predict the large-scale seasonal tropical circulation and precipitation for as long as the SSTs can be predicted
2. Tropics act as a source of e.g. Rossby waves which can propagate into the distant extra-tropics → *Predictability in the Midst of Chaos*

The extent to which this high predictability in the atmosphere can be realized depends on our ability to predict the SST itself.

Assessing the skill of SEAS5



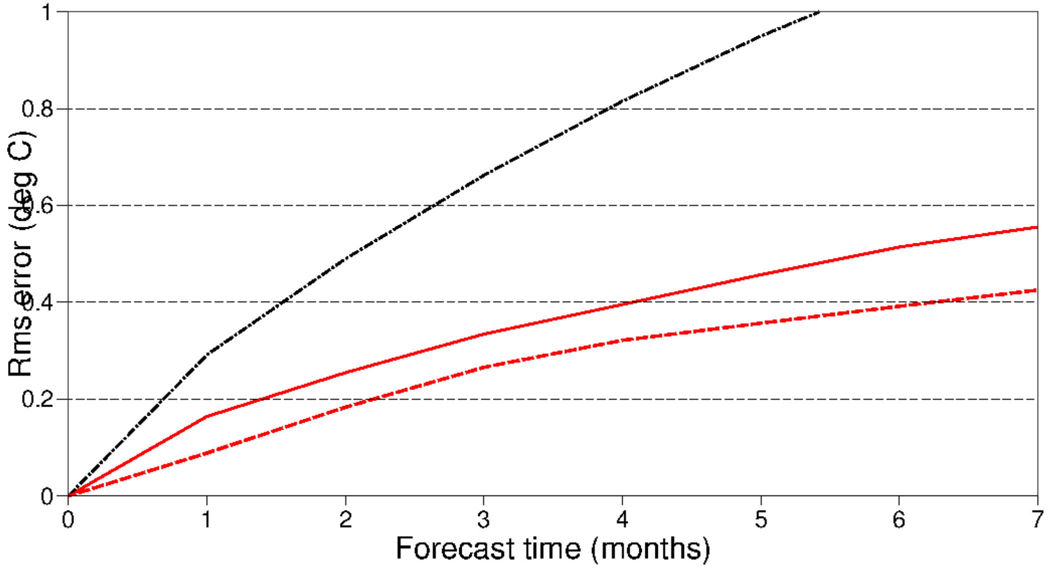


Nino3.4 SST drift

Local SST bias is a function of

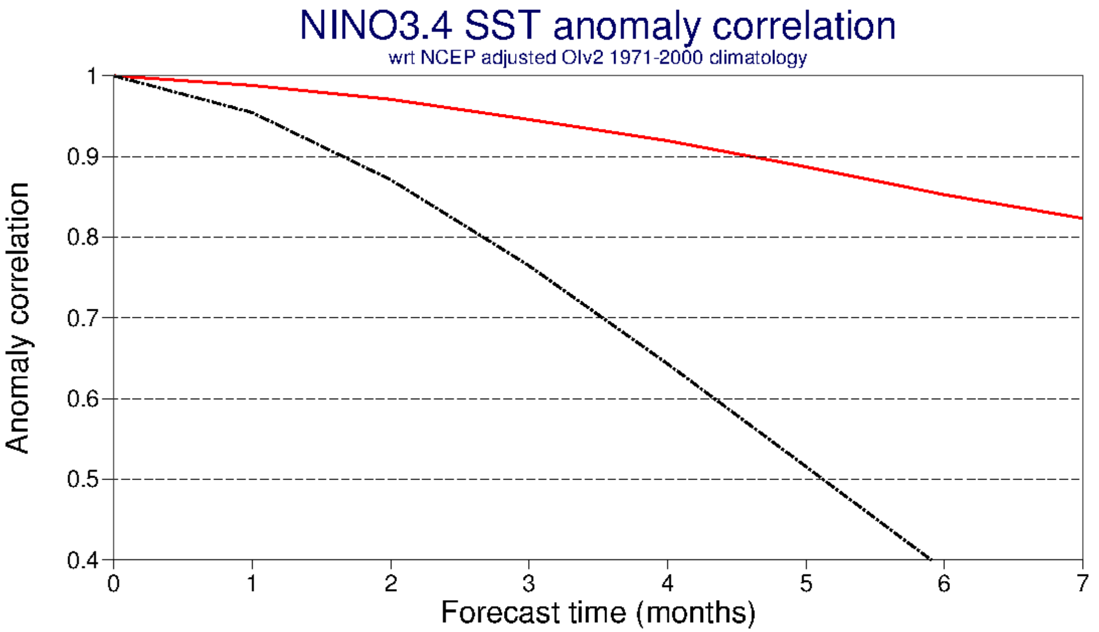
- **lead time**
- **season**

System 5 (Nov 2017 onwards)
System 4 (Nov 2011 - Oct 2017)
System 3 (March 2007 - Nov 2011)



Nino3.4 SST forecast error and ensemble spread: lead-time dependence (all seasons)

RMSE forecast error
Ensemble spread
RMSE of persistence



In a *perfect model*:

RMSE ≃ ensemble spread

In *SEAS5*:

actual RMS errors > "perfect model" errors

or ensemble spread too small (overconfidence)

forecast anomaly correlation
persistence anomaly correlation

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
- Time-mean RMSE of the ensemble mean equals the time-mean ensemble spread

→ spread is an indication of perfect model error

From Palmer et al. (2006)

Box A The relationship between spread and ensemble mean RMS error in a perfect ensemble

In a perfect ensemble, i.e. a perfect sampling of the underlying probability distribution of truth, then, over a large number of ensemble forecasts, the statistical properties of the true value X_T of X are identical to the statistical properties of a member of the ensemble, X_e (when that member is removed from the ensemble). For the following analysis of spread and skill, we assume that the ensemble size N is sufficiently large that removing one member from the ensemble does not materially affect the results. Hence, for example, the mean squared distance of the J-th member $X_e(J)$ from the ensemble mean $\langle X_e \rangle$ is identical to the mean squared error of the ensemble mean

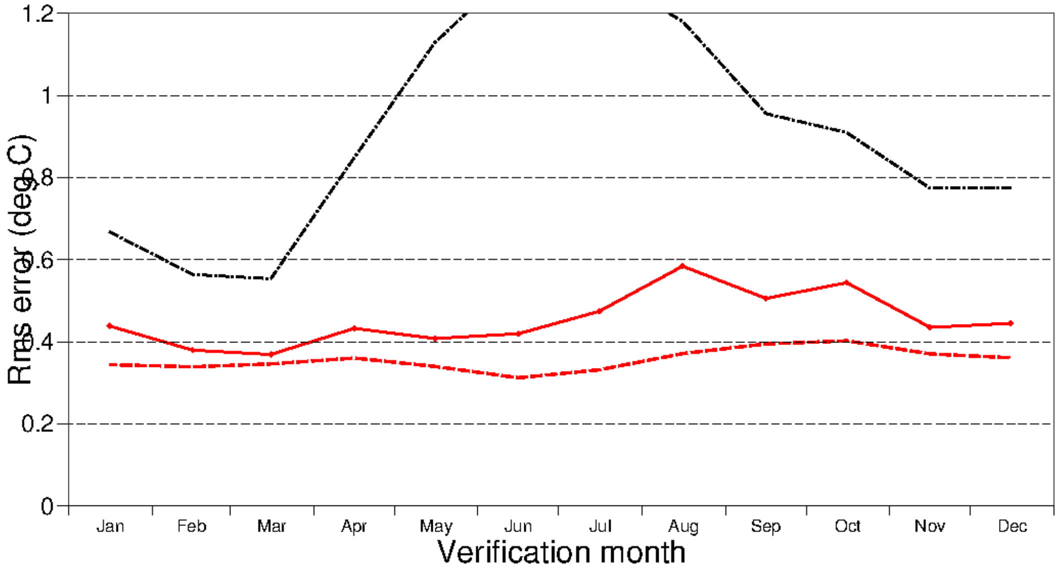
$$\overline{\left\| X_e(J) - \langle X_e \rangle \right\|^2} = \overline{\left\| X_T - \langle X_e \rangle \right\|^2} \qquad \text{(A1)}$$

where $\langle \ldots \rangle$ denotes the expectation value with respect to a particular ensemble forecast, and $\overline{\ldots}$ denotes an average over many such ensemble forecasts. Equation (A1) holds for any J and it can be applied to a scalar quantity X or to a vector $\boldsymbol{X}$. In the latter case, $\| \ldots \|$ should be understood as the Root Mean Square (RMS) or the Euclidean norm. Taking the expectation $\langle \ldots \rangle$ of Equation (A1) yields

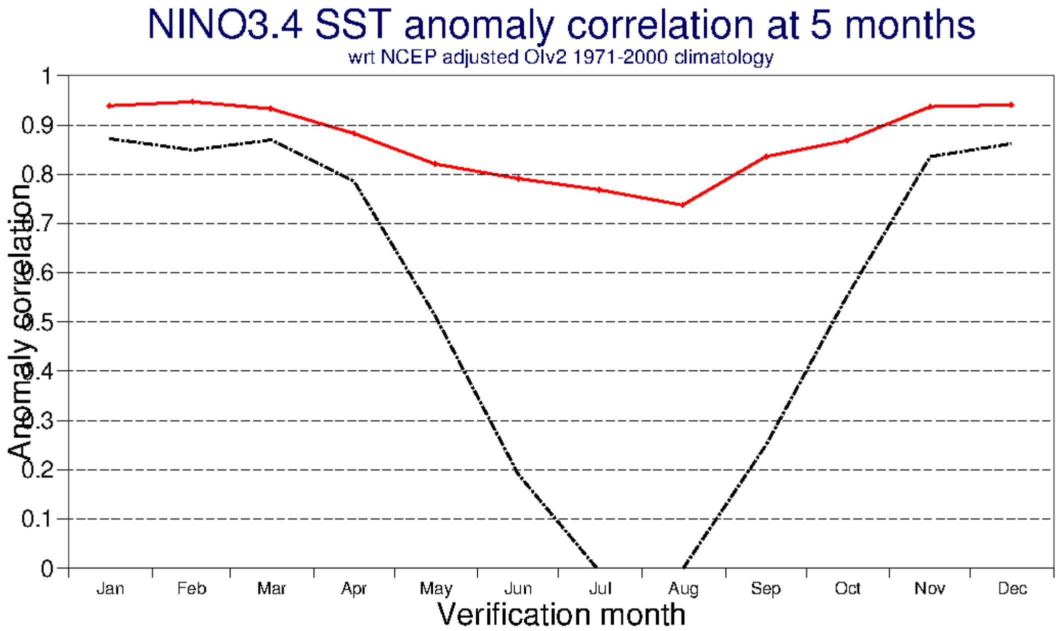
$$\overline{\left\langle \left\| X_e - \langle X_e \rangle \right\|^2 \right\rangle} = \overline{\left\| X_T - \langle X_e \rangle \right\|^2} \qquad \text{(A2)}$$

Equation (A2) implies that the time-mean ensemble spread about the ensemble-mean forecast, should equal the time-mean RMS error of the ensemble-mean forecast.

Nino3.4 SST forecast error and ensemble spread: seasonal dependence (lead-time: 5 months)



RMSE forecast error
Ensemble spread
RMSE of persistence



Spring predictability barrier

forecast anomaly correlation
persistence anomaly correlation

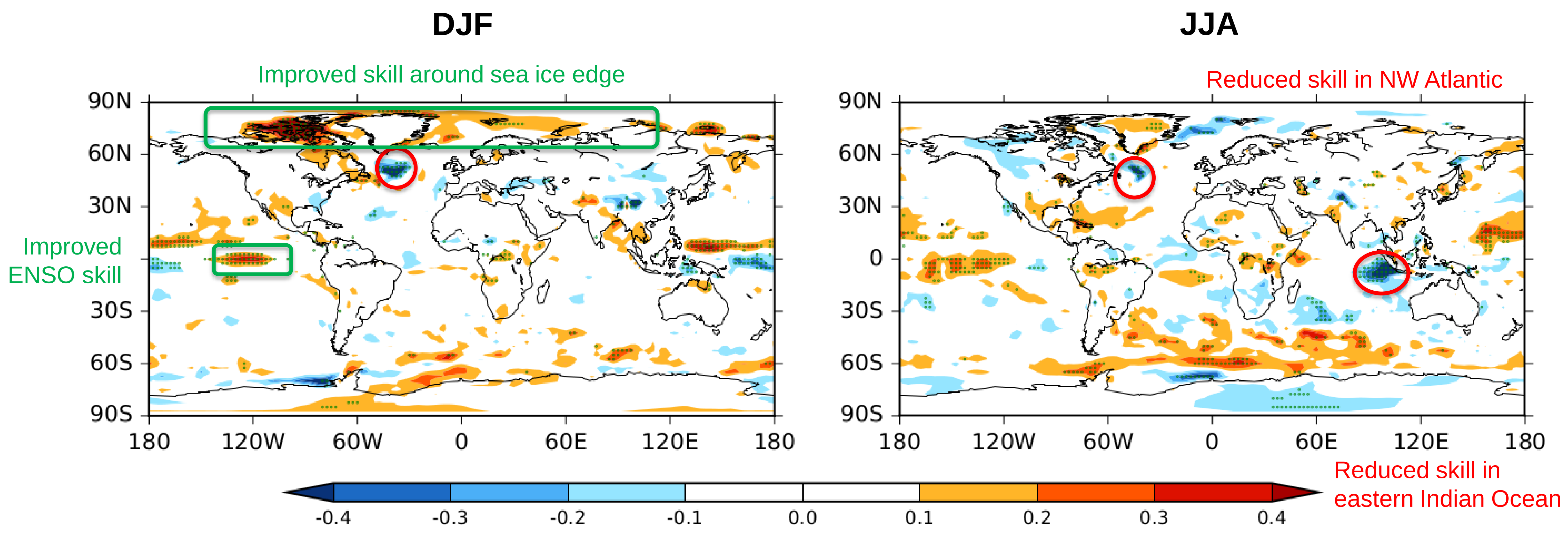
System 4 vs SEAS5

	System 4 Introduced: 2011	SEAS5 Introduced: 2017
Atmosphere	Cycle 36r4 T_L255 L91	Cycle 43r1 T_{Co}319 L91
Ocean	NEMO v3.0 ORCA 1.0-L42	NEMO v3.4 ORCA 0.25-L75
Sea ice model	Sampled climatology	LIM2
Atm. initial conditions	ERA-Interim/Ops	ERA-Interim/Ops
Ocean and sea ice initial conditions	OCEAN4	OCEAN5

SEAS5 documentation: Johnson et al. (2019)

Global 2m temperature skill differences

$CRPSS_{SEAS5} - CRPSS_{SEAS4}$

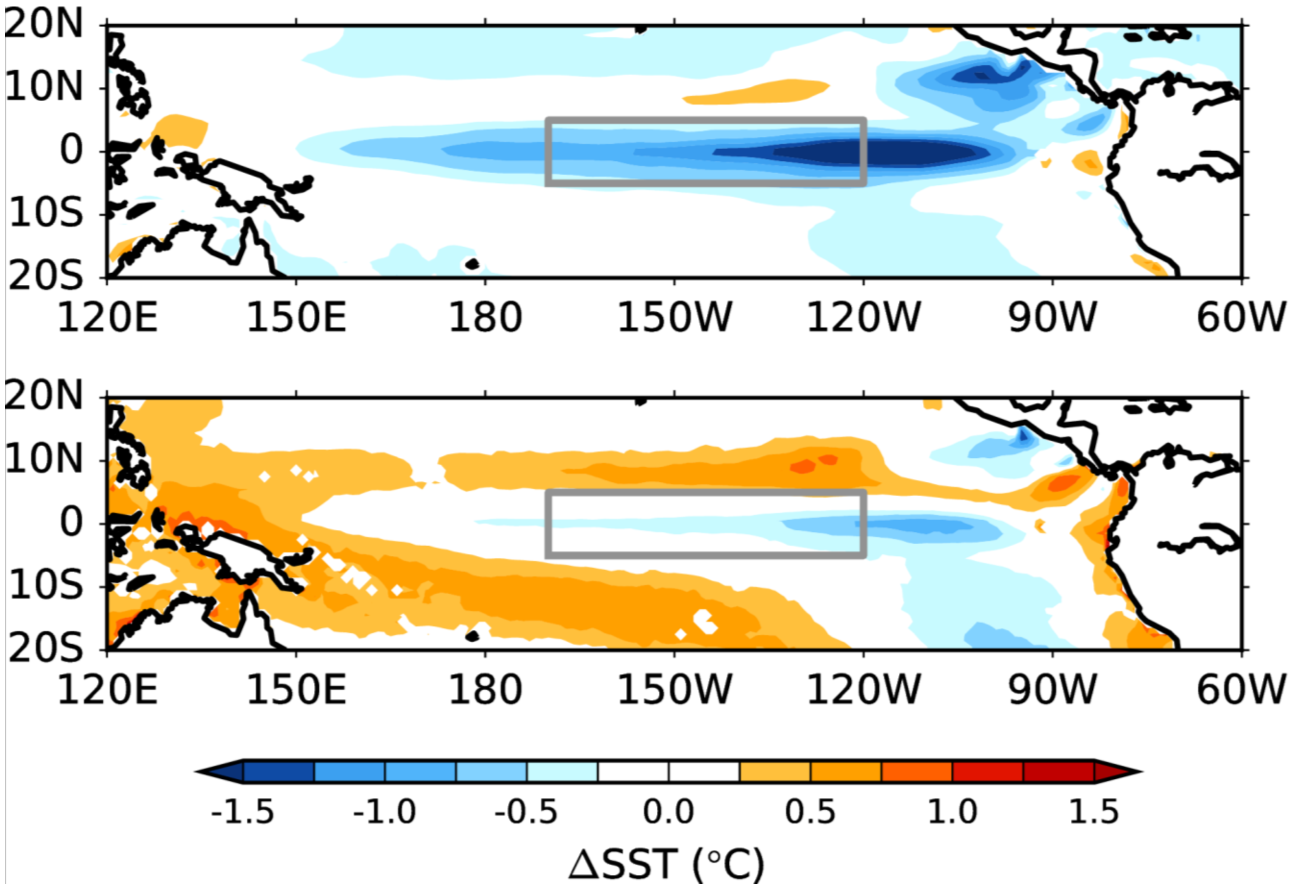


All plots show seasonal means at one month forecast lead (i.e. month 2 to 4 of the forecast) and using 25 ensemble members, unless stated otherwise

Nino3.4 SST bias in DJF

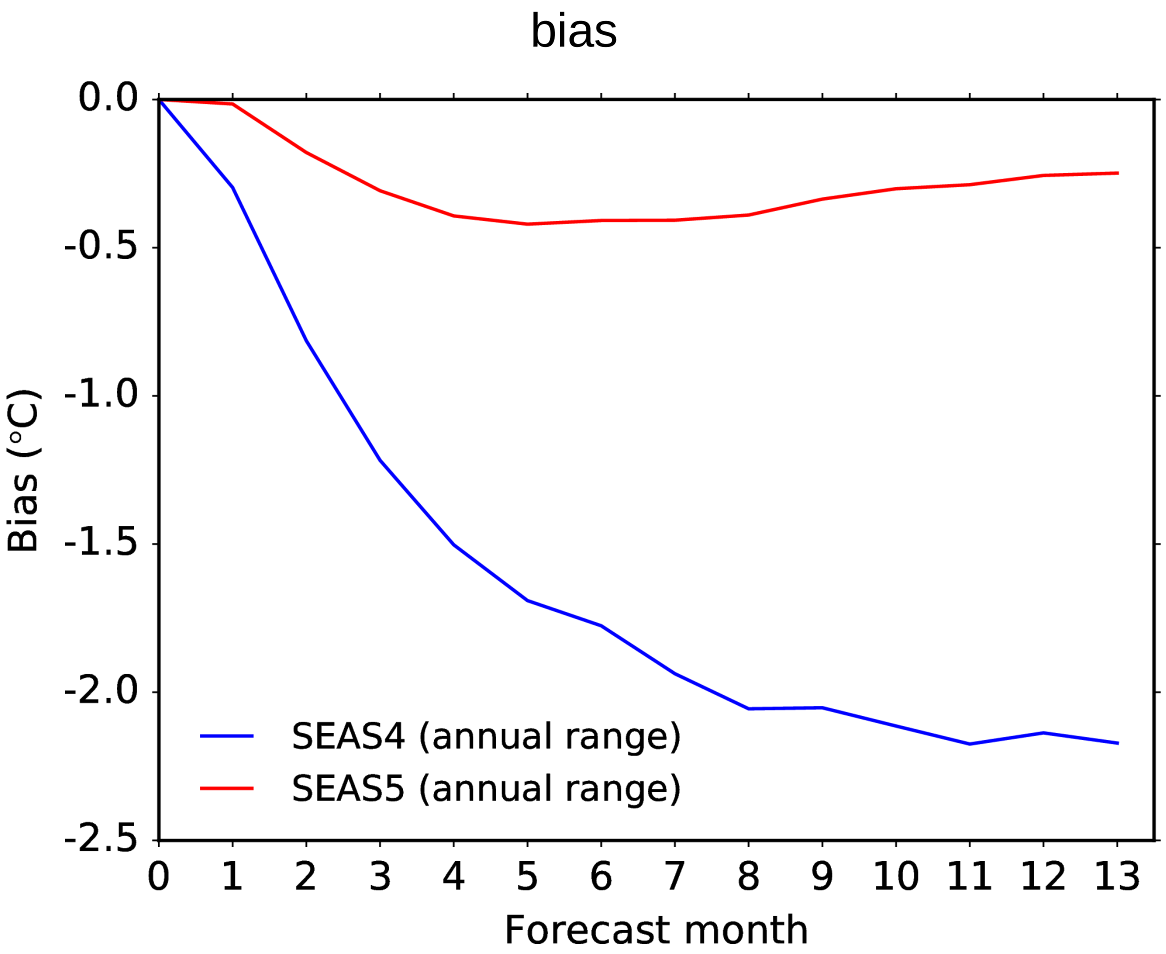
System 4

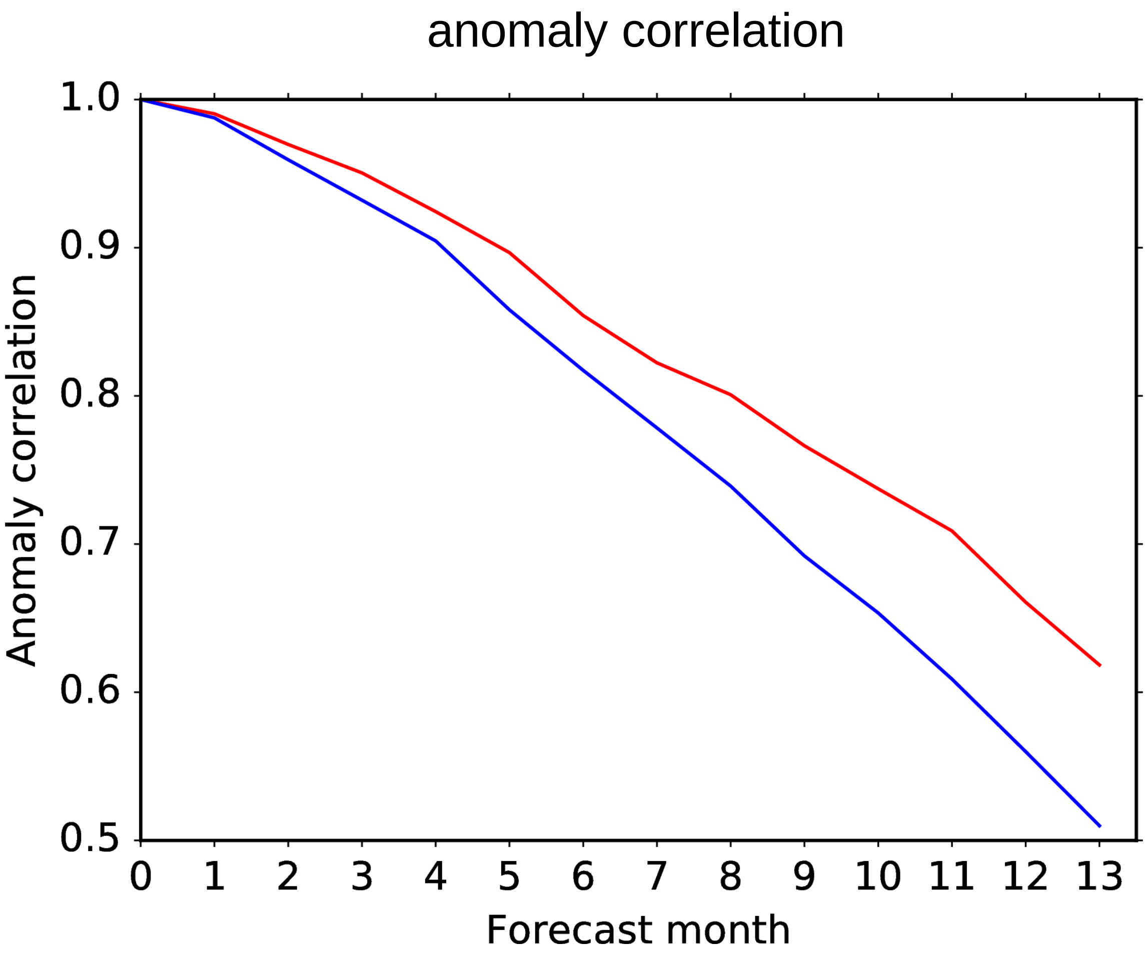
SEAS5



Large improvement in equatorial cold tongue bias

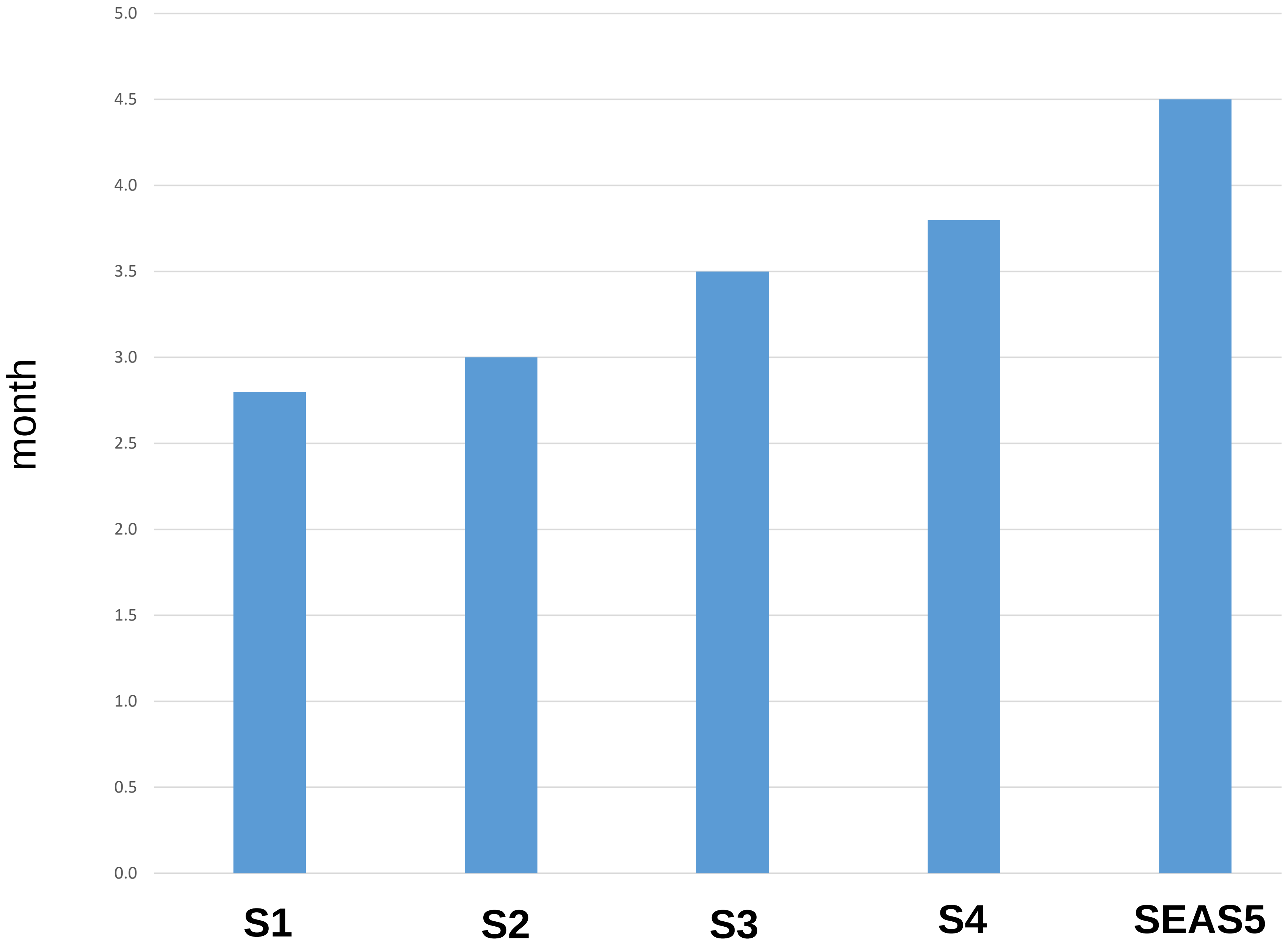
Nino3.4 SST skill





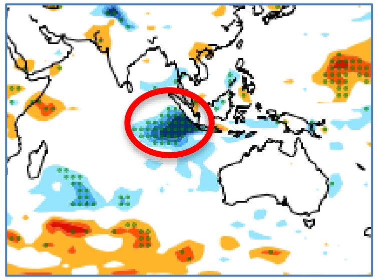
Improvement in Niño 3.4 skill, particularly at longer lead times

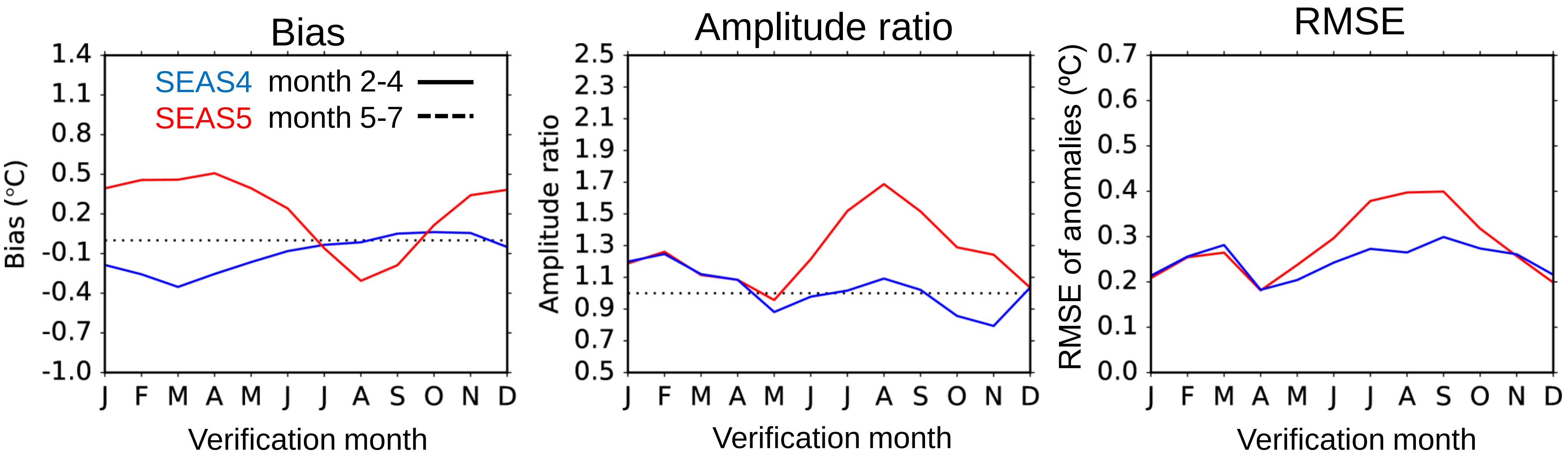
Forecast lead time when ACC (NINO3.4) < 0.9



Eastern Indian Ocean

ΔCRPSS



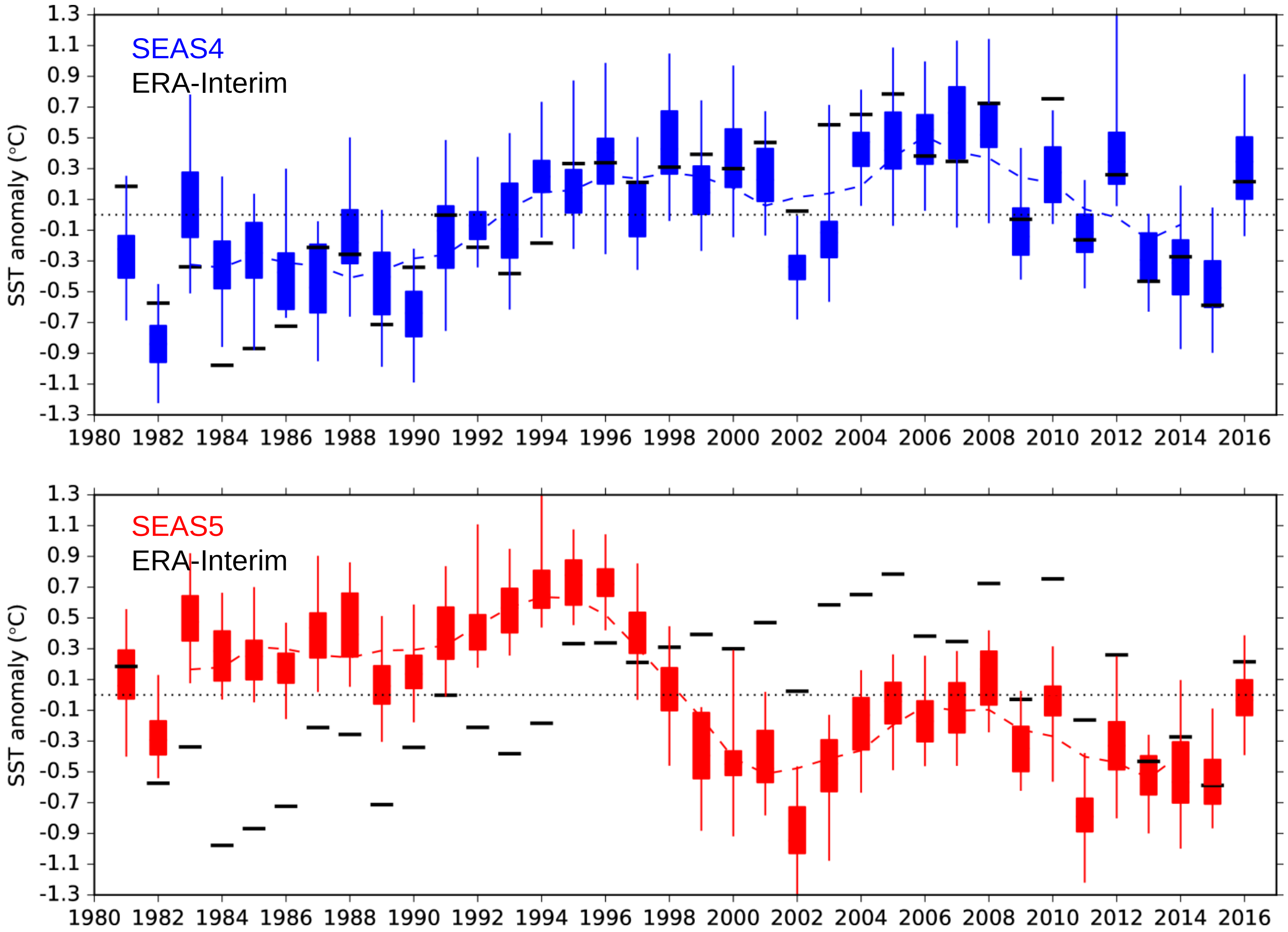


Cold SST anomalies in the eastern Indian Ocean are too large, too variable and too frequent

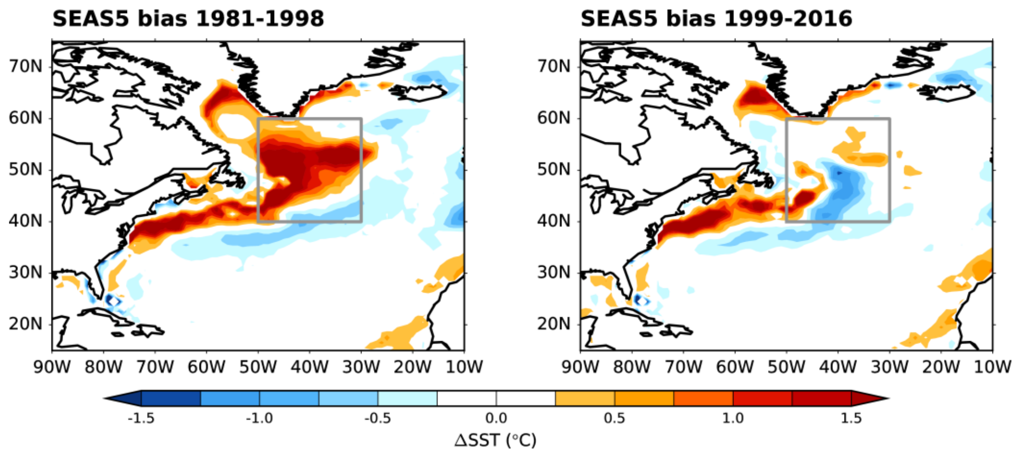
Results in large errors in skill in the eastern box of Indian Ocean Dipole index

Northwest Atlantic

DJF SST anomalies in Northwest Atlantic



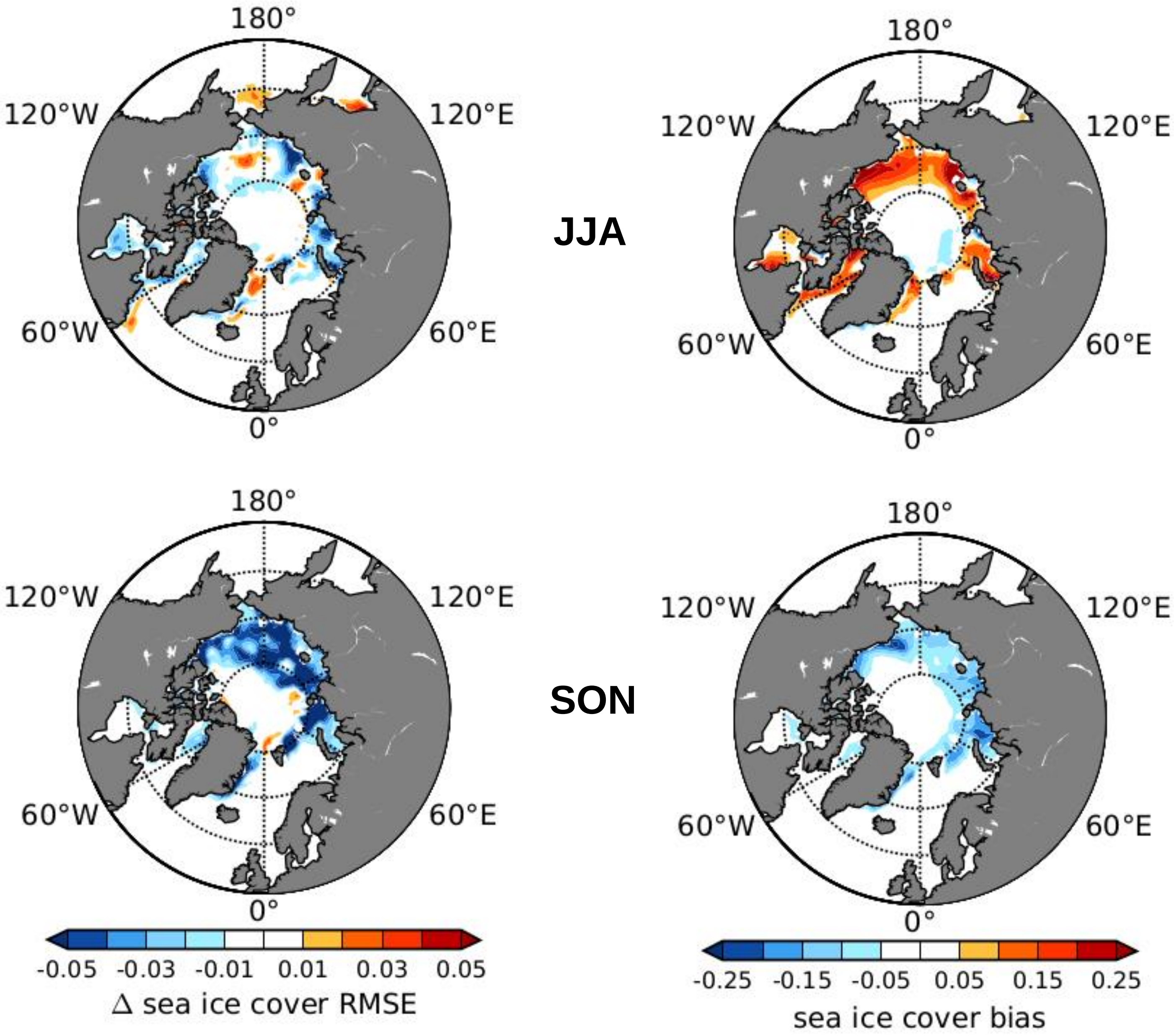
SEAS5 DJF SST bias



- **Poor representation of the decadal variability in the Northwest Atlantic**
- **Related to increased resolution of new ocean analysis system**

Artic sea ice

SEAS5 – SEAS4 RMSE **SEAS5 bias**



SEAS4 used a simple empirical scheme that only captured the trend but not the interannual variability

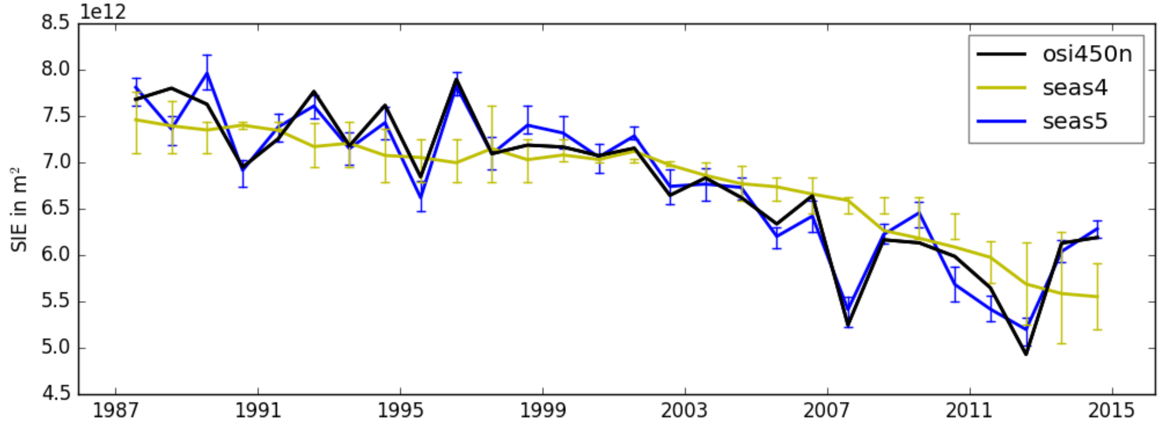
Adding LIM2 improves skill in predicting sea ice but introduces sea ice biases

→ Largest biases in summer

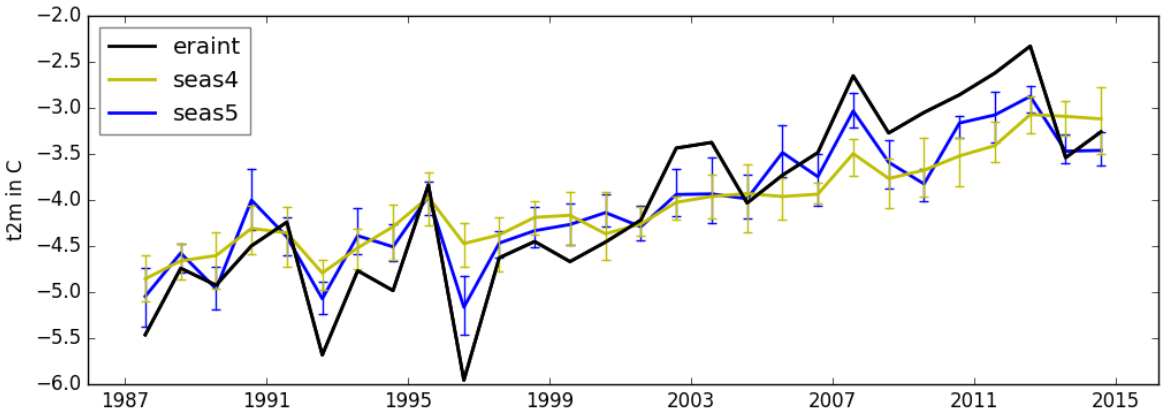
→ Biggest skill improvements in autumn

Artic sea ice

ASO mean sea ice extent north of 70⁰ N
July start - one month lead

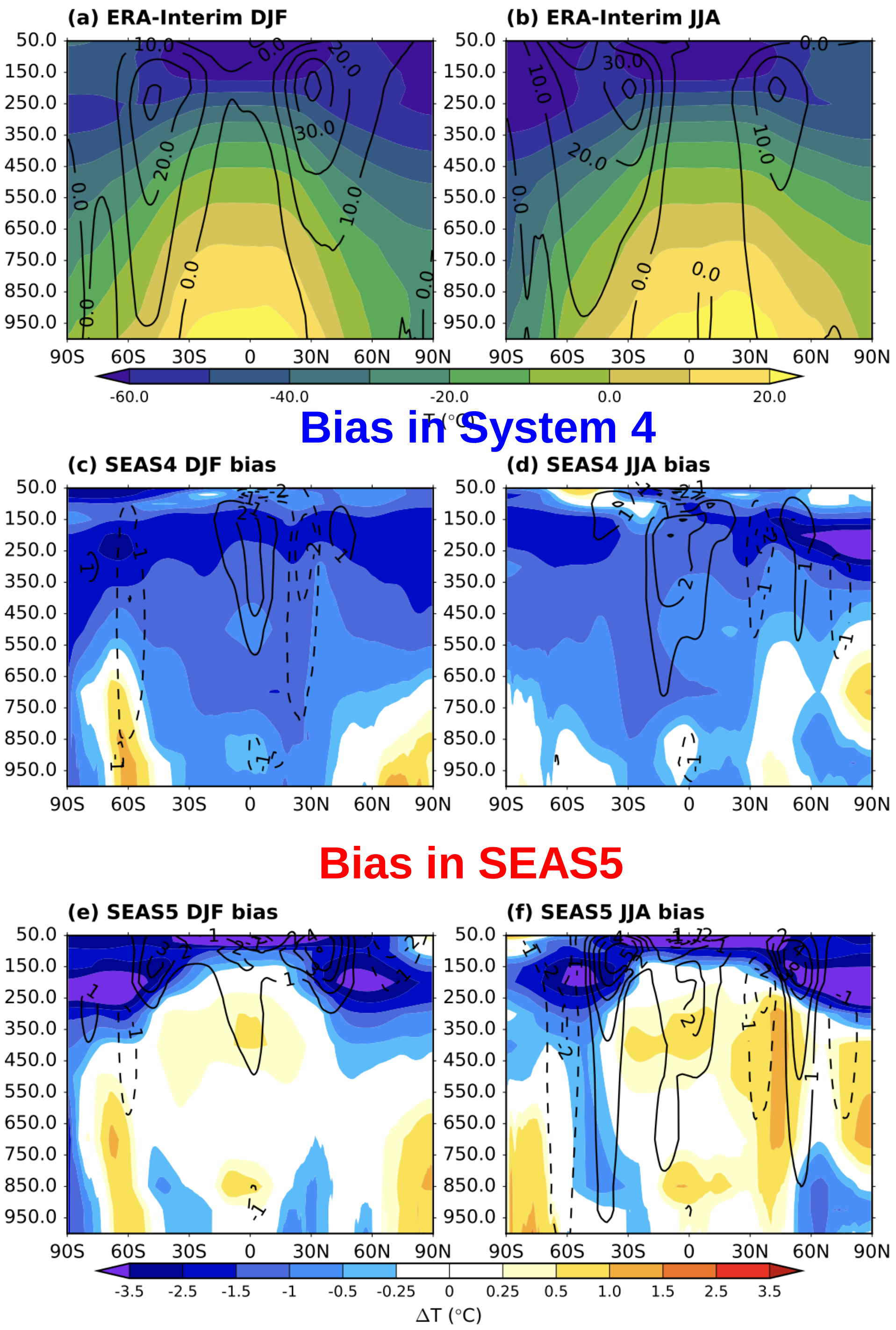


ASO mean 2 m temperature north of 70⁰ N
July start - one month lead



Prognostic sea ice model introduces interannual variability in arctic sea ice extent

Increased skill in T2m north of 70⁰ N associated with improved sea ice prediction



Zonally averaged profiles of

- **temperature (colour)**
- **zonal wind (contours)**

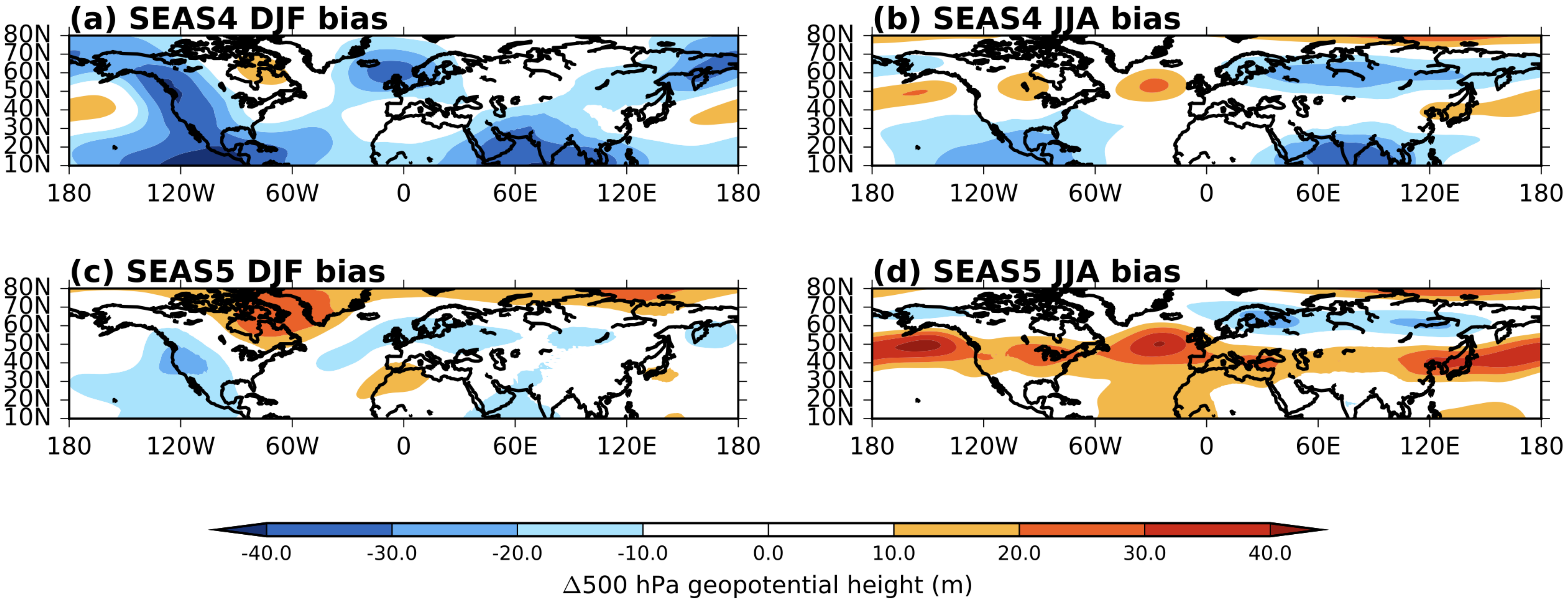
Largely reduced tropospheric temperature biases in DJF

Too strong and poleward displaced jets, especially in JJA

Strong cold tropopause biases

From Johnson et al. (2019)

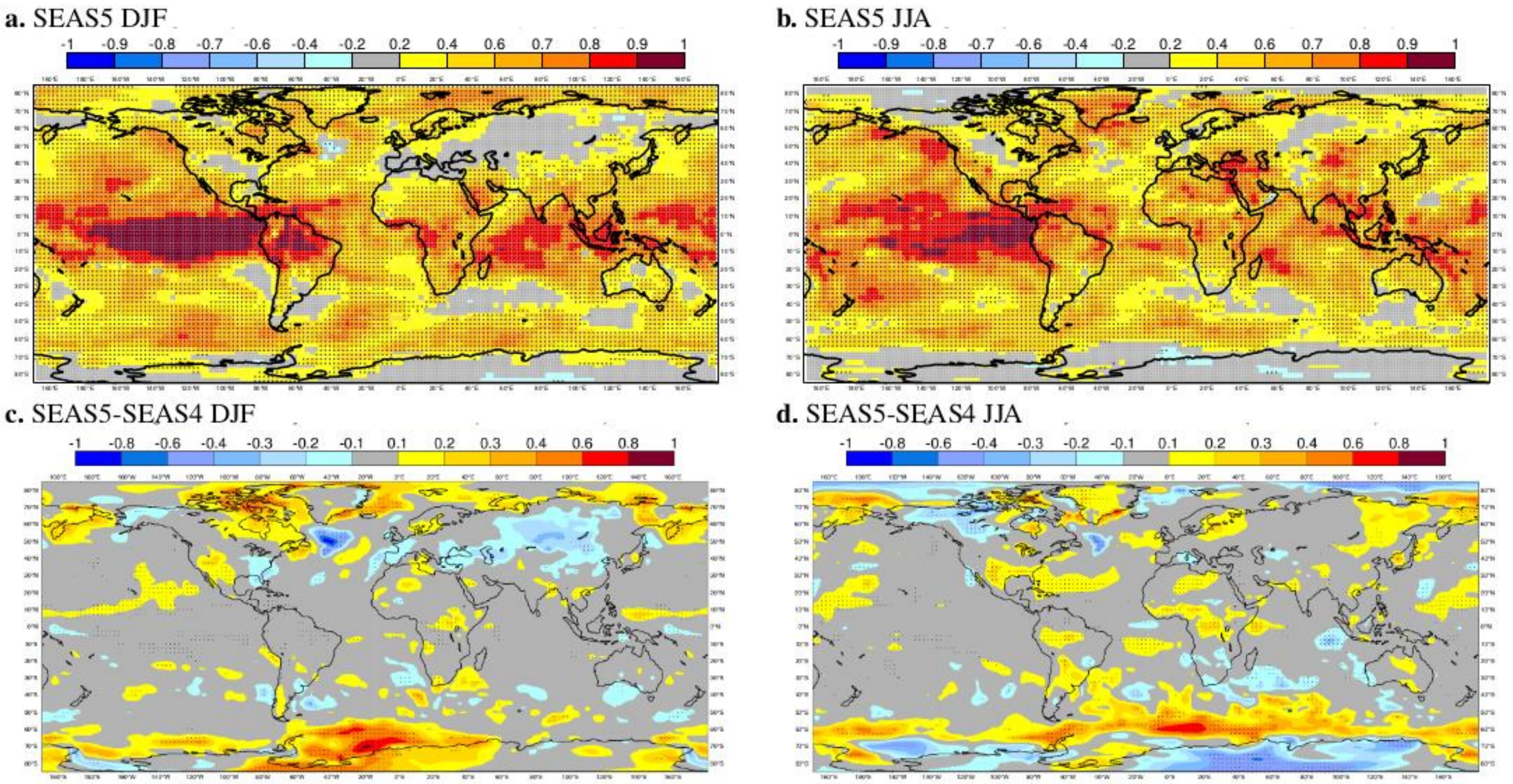
Bias of geopotential height at 500 hPa



Mean-state changes in line with tropopause warming in SEAS5
Strong jet biases in JJA

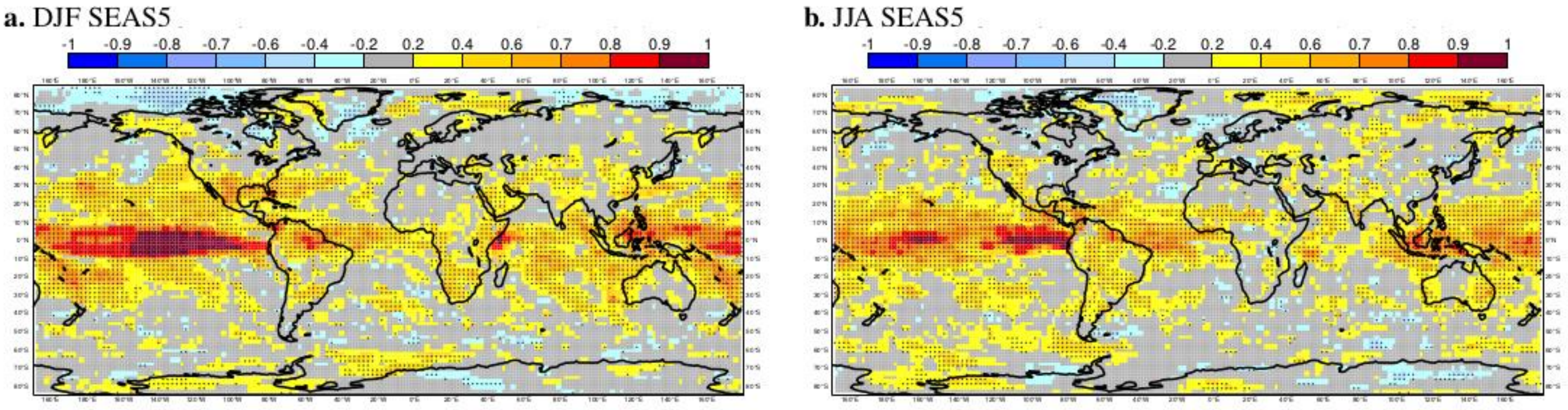
From Johnson et al. (2019)

Ensemble mean correlation skill for 2m temperature



From Johnson et al. (2019)

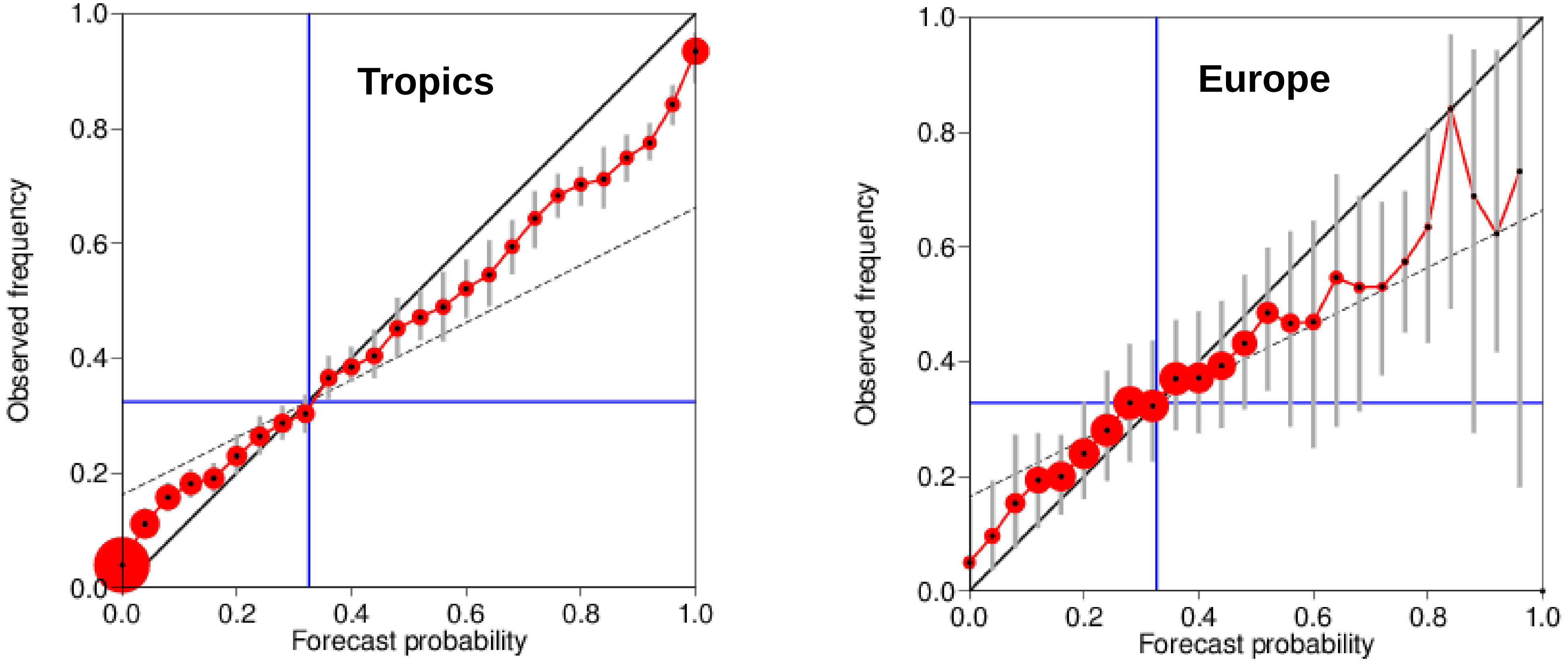
Ensemble mean correlation skill for precipitation



From Johnson et al. (2019)

Reliability diagrams

2m temperature upper tercile events in DJF



Overconfident forecasts

From Johnson et al. (2019)

Hindcasts of the NAO in DJF

System 4
SEAS5
ERA-I

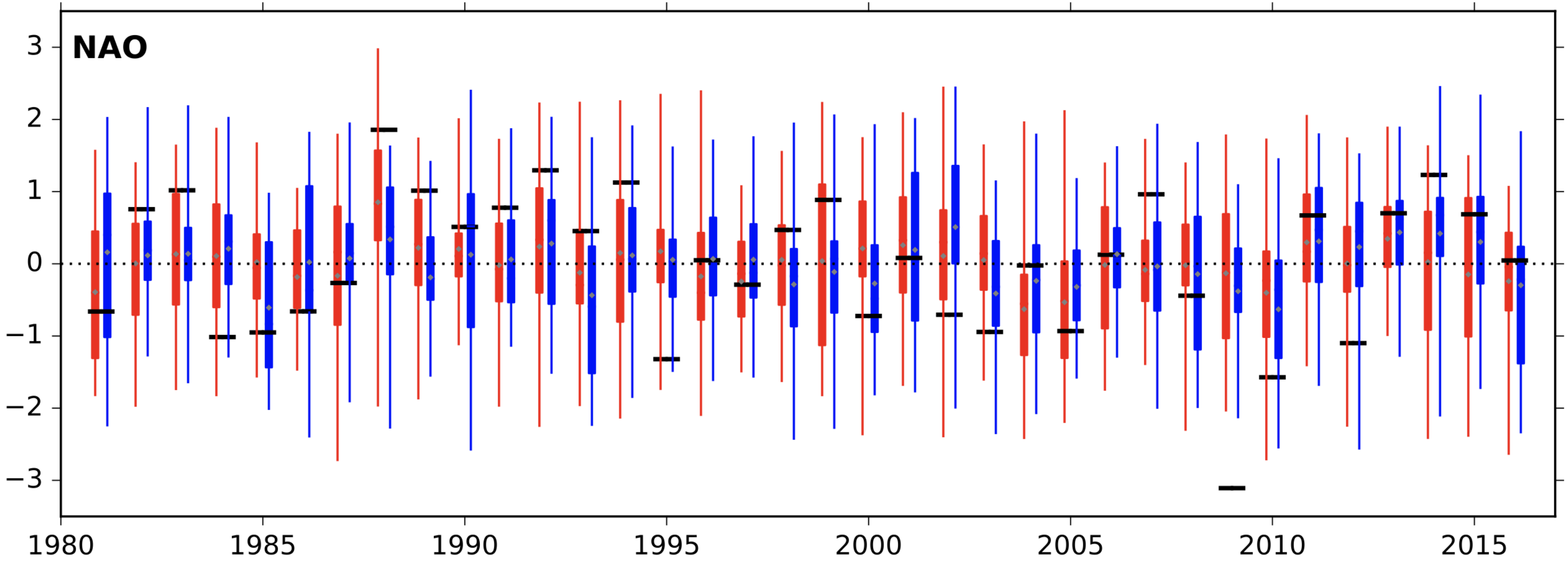
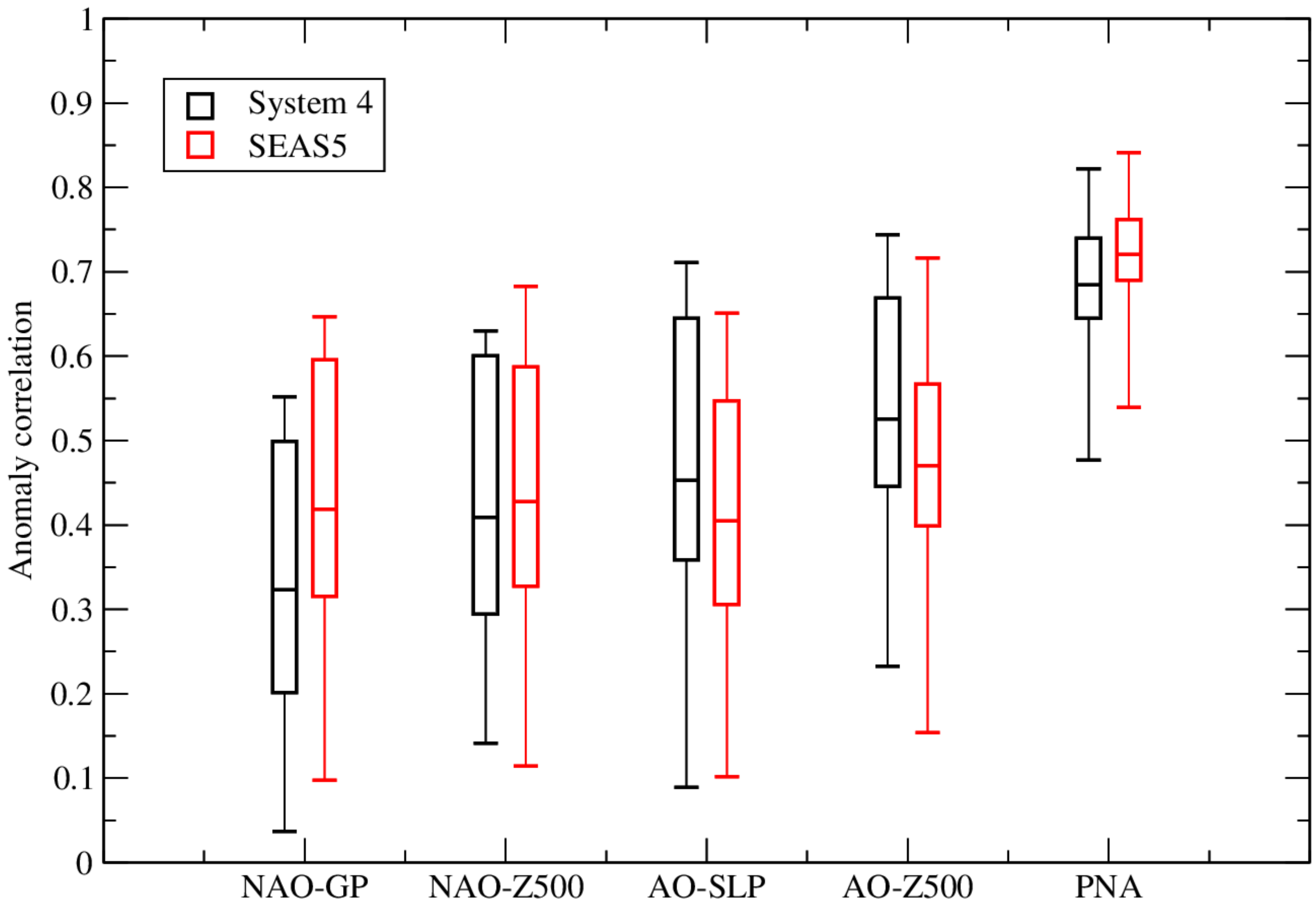


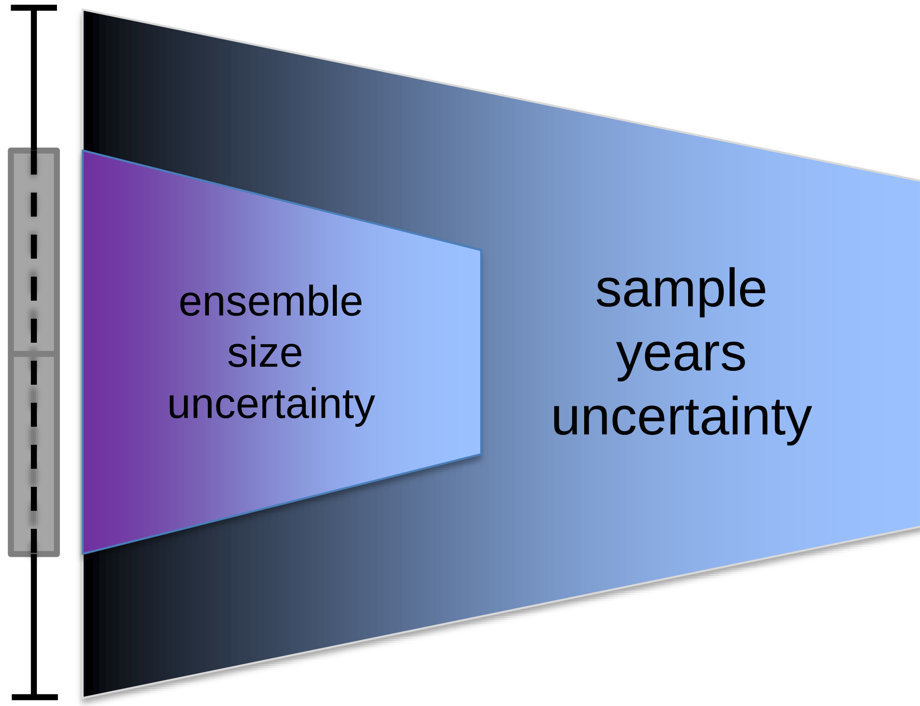
Figure 11. Time series of a DJF NAO index derived from projecting the re-forecast 500 hPa geopotential height onto the first EOF of ERA-Interim 500 geopotential height in the North Atlantic. Quartiles, minimum and maximum of the SEAS4 25 member ensemble are shown in blue, while the SEAS5 25 member ensemble is shown in red and ERA-Interim reanalysis is shown in the black bars. Forecasts were initialised in November, and the year shown is the year the ensemble was initialised. The grey diamonds indicate the ensemble mean. Anomaly correlation values for the ensemble mean are 0.45 for SEAS4 and 0.44 for SEAS5. The 95% confidence interval for sampling error over years is 0.12 to 0.67 for SEAS5.

From Johnson et al. (2019)

Hindcast skill of the NAO, AO and PNA in DJF

Nov start dates 1981-2016, 51 ensemble members

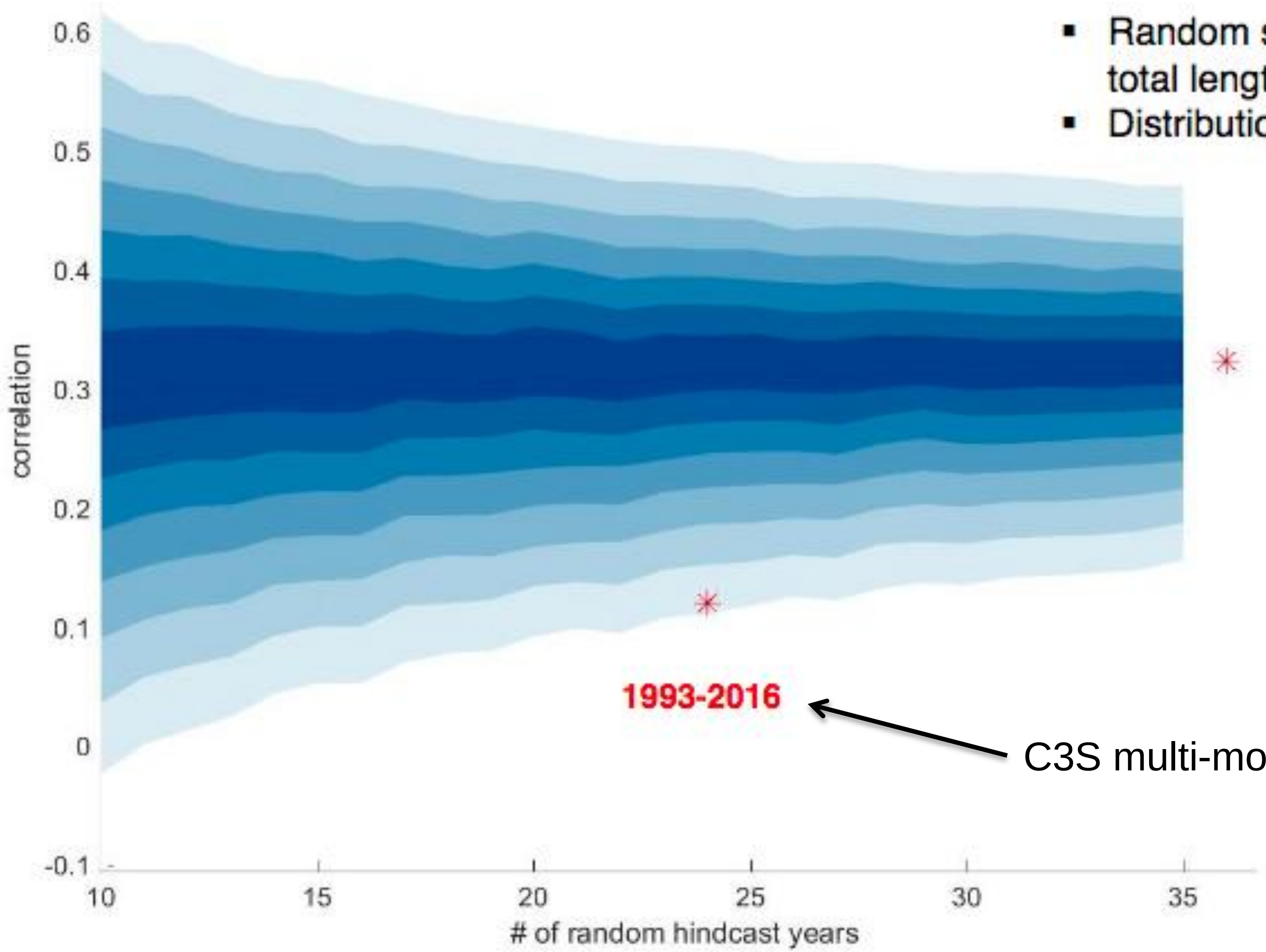




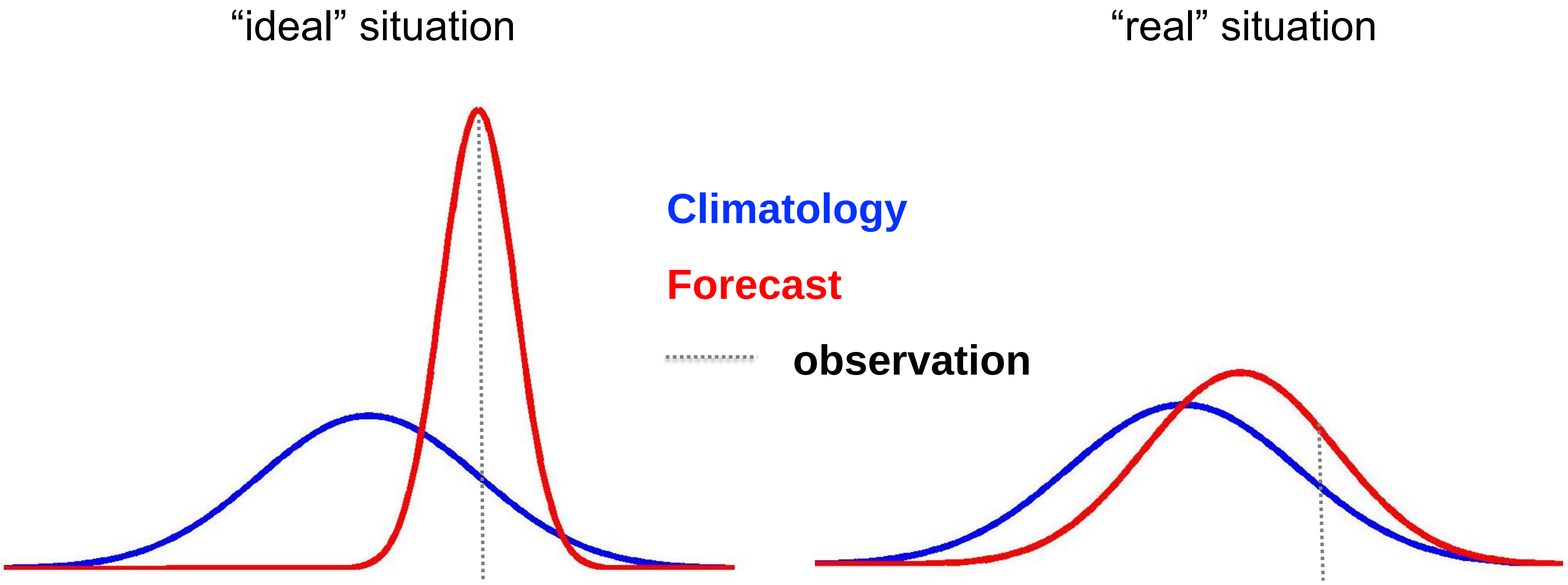
Two types of sampling uncertainty

NAO hindcast skill in SEAS5 – how (un)lucky are we?

- Random samples of hindcast years with total length between 10 and 34 years
- Distribution of correlation



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



Seasonal forecasts of the winter NAO

THE TIMES

Science

April 2014

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

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

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

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.

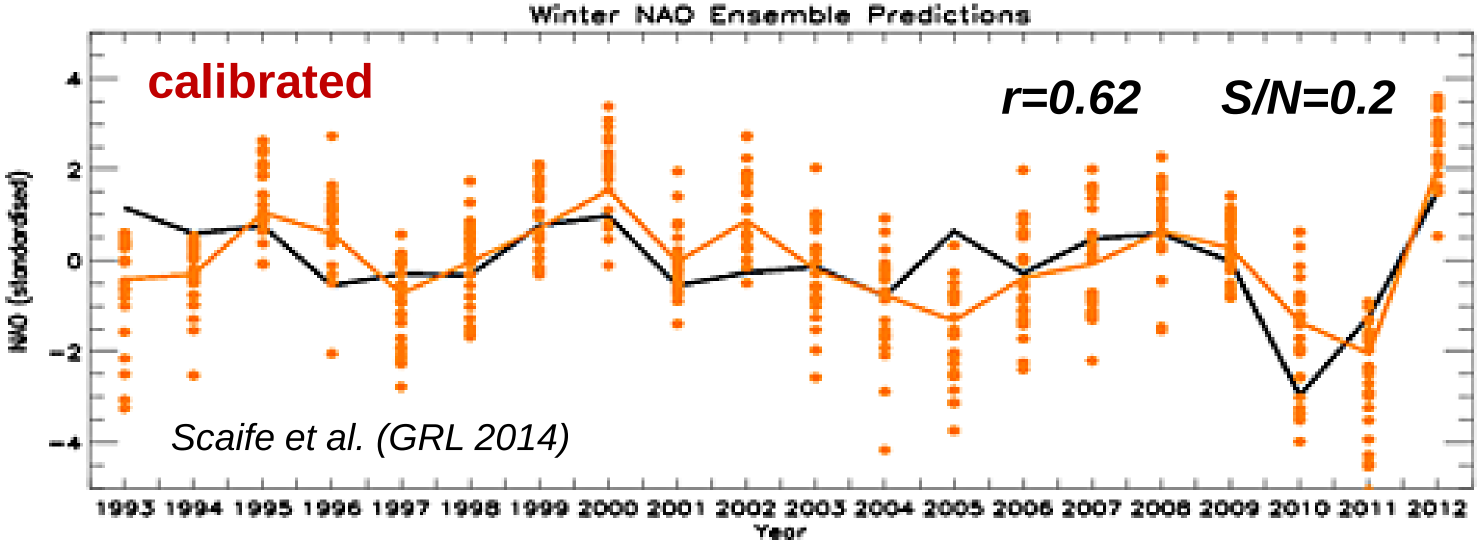
The breakthrough may have a substantial impact on the economy, allowing power companies and wind farms to anticipate energy demands while airports and councils can estimate how much grit and anti-freeze is likely to be required.

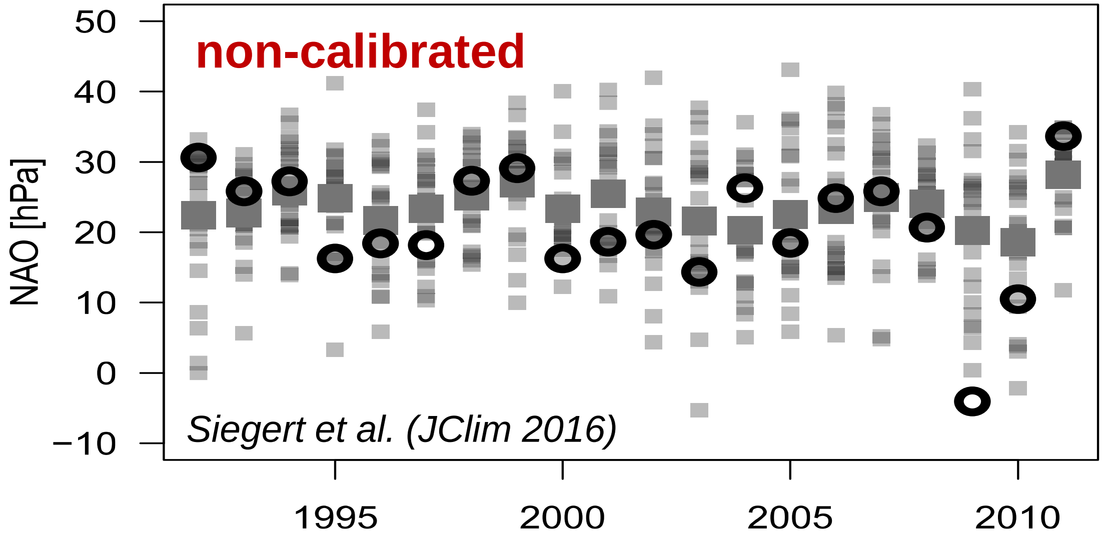
Print

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$_{noise}$***

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

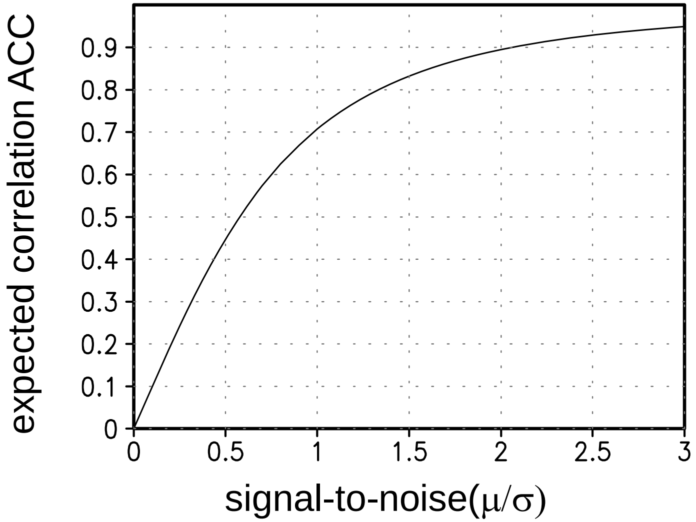
ensemble mean variance
*→ "**signal**"*

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

*variance of ensemble members about ensemble mean → "**noise**"*

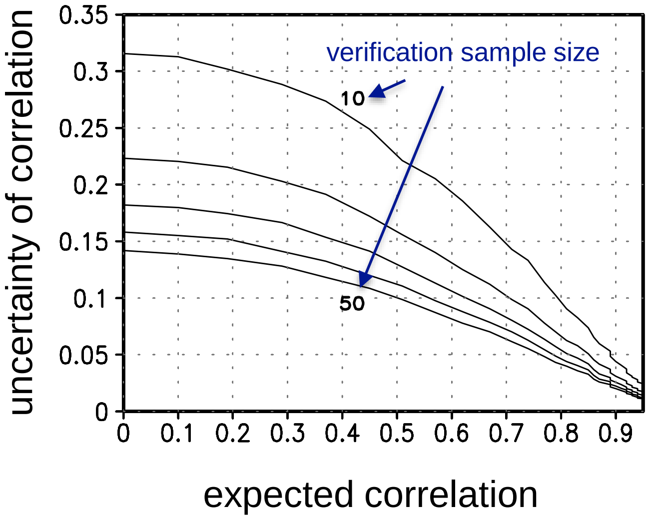
Correlation skill and signal-to-noise (S/N) ratio

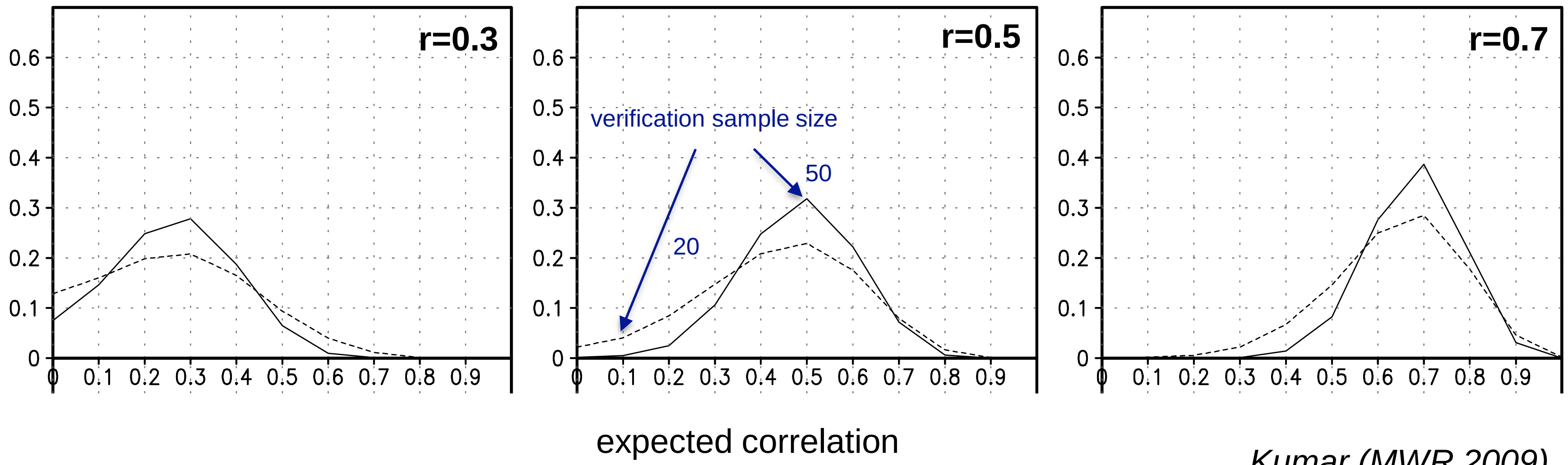
The *expected value* for various measures of skill for seasonal climate predictions is determined by the S/N ratio.



"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 Had

Abstract S

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

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

observed PC_{obs} … estimated from explained variance = $\boldsymbol{r}^2(obs, ensmean)$

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

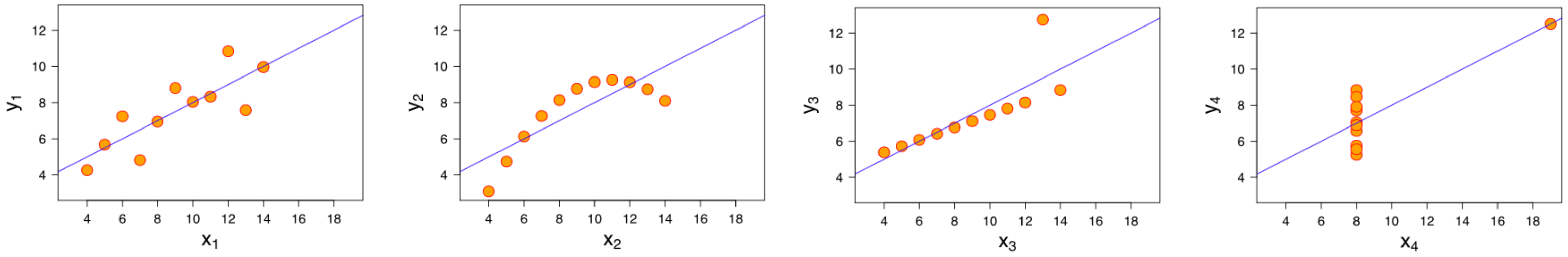
Properties of a perfect model ensemble

- Time-averaged ensemble spread *approx. equals* RMSE of ensemble mean forecast
- Correlation skill (aka potential skill): correlation(ensemble mean, ensemble member)
- Observed correlation ≤ perfect model correlation
- RPC of a perfect ensemble *approx. equals* 1

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

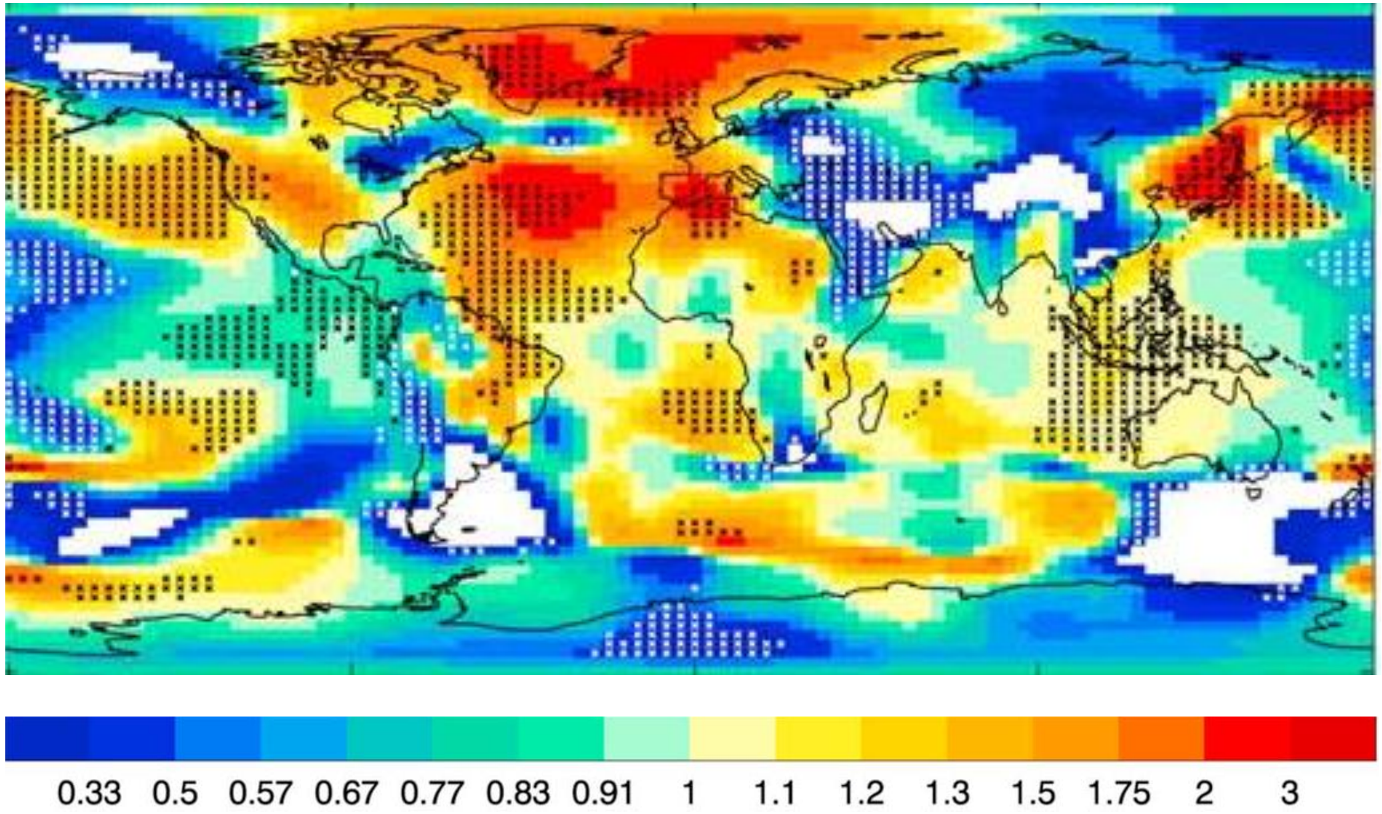


Normally distributed, “well behaved” | Not normally distributed, non-linear relationship | Perfect linear relationship, Except for one outlier | No relationship with one outlier

Anscombe (Amer. Statist. 1973)

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)

The real world seems to have higher predictability than the model

Thank you!

References

Anscombe, F.J. 1973. Graphs in statistical analysis. *The American Statistician* 27, **1**, 17-21. See also https://en.wikipedia.org/wiki/Anscombe%27s_quartet.

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

Johnson, S.J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., and Monge-Sanz, B. (2019). SEAS5: The new ECMWF seasonal forecast system, *Geosci. Model Dev. Discuss.*, **12**, 1087–1117, doi:10.5194/gmd-2018-228

Kumar, A., A.G. Barnston and M.P. Hoerling. 2001. Seasonal predictions, probabilistic verifications, and ensemble size. *J. Clim.,* **14**, 1671-1676.

Kumar, A. 2009. Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Mon. Weather Rev.*, **137**, 2622-2631.

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.

O'Reilly, C.H., A. Weisheimer, T. Woollings, L. Gray and D. MacLeod. 2018b. The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj3413.

Palmer, T., R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher and L. Smith (2006). Ensemble prediction: A pedagogical perspective. *ECMWF Newsletter*, **106**, 10-17.

References

Scaife, A.A., et al. (2014). Skilful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514-2519, doi:10.1002/2014GL059637.

Scaife, A.A. and D. Smith (2018). A signal-to-noise paradox in climate science. npj Climate and Atmospheric Science, **1**, 28.

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

Siegert, S., D.B. Stephenson and P.G. Sansom (2016). A Bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is NAO predictability? *J. Clim.*, **29**, 995-1012, doi: 10.1175/JCLI-D-15-0196.1

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.

Stockdale, T., S. Johnson, L. Ferranti, M. Balmaseda and S. Briceag. 2018. ECMWF's new long-range forecasting system SEAS5. *ECMWF Newsletter*, **154**, 15-20.

Weisheimer, A., N. Schaller, C. O'Reilly, D. MacLeod and T.N. Palmer. 2017. Atmospheric seasonal forecasts of the 20th Century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation and their potential value for extreme event attribution. *Q. J. R. Meteorol. Soc.,* **143**, 917-926, doi:10.1002/qj.2976.

Weisheimer, A., D. Decremer, D. MacLeod, C. O'Reilly, T. Stockdale, S. Johnson and T.N. Palmer (2018). How confident are predictability estimates of the winter North Atlantic Oscillation? *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.3446

Weisheimer, A., D. Befort, D. MacLeod, T.N. Palmer, C. O'Reilly and K. Strommen (2020). Seasonal forecasts of the 20th Century *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-19-0019.1