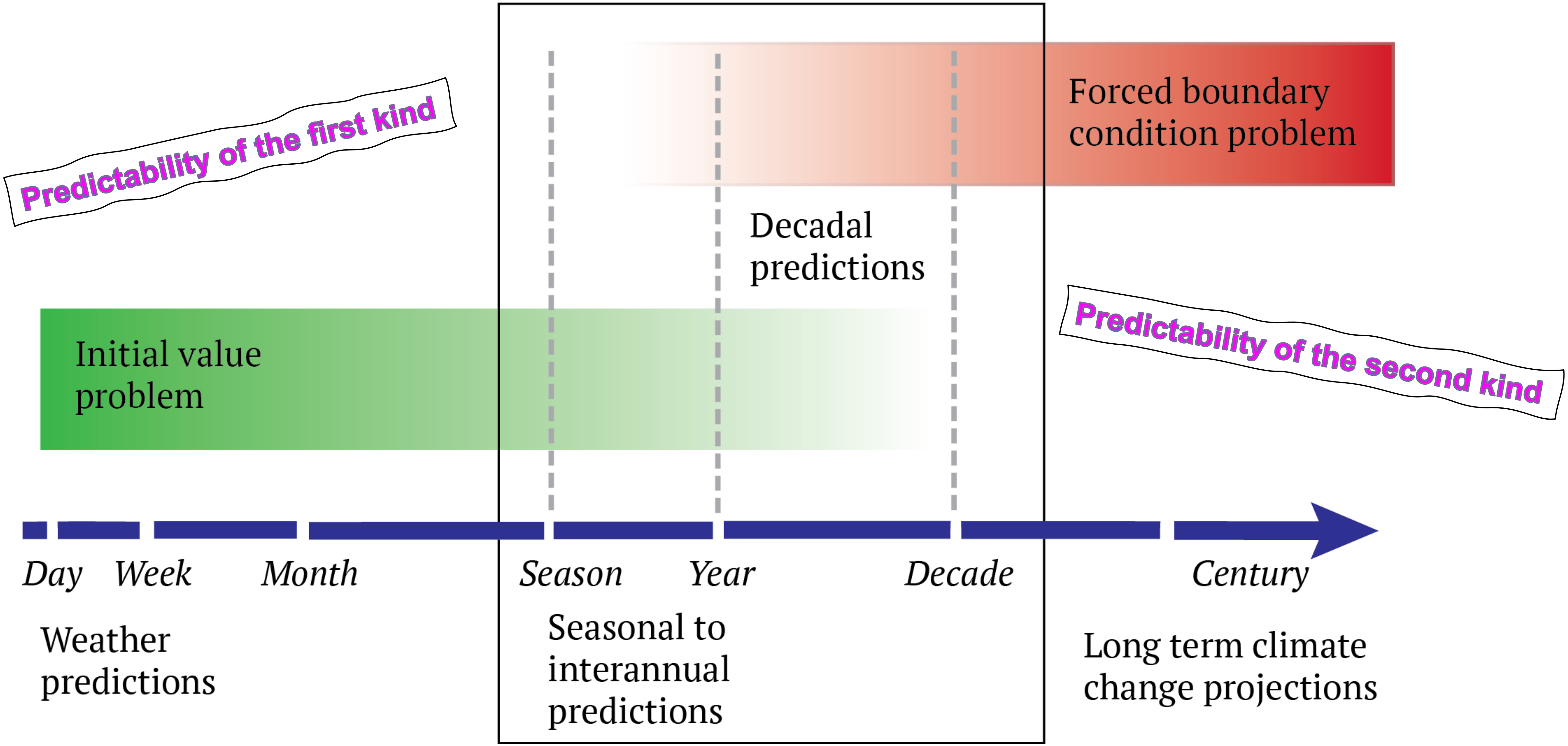
Seasonal predictability: How confident can we be?

Antje Weisheimer

ECMWF, Earth System Predictability Section

University of Oxford, Atmospheric, Oceanic and Planetary Physics

What are seasonal forecasts?



IPCC AR5

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)

El Niño Southern Oscillation – a source of predictability on seasonal timescales

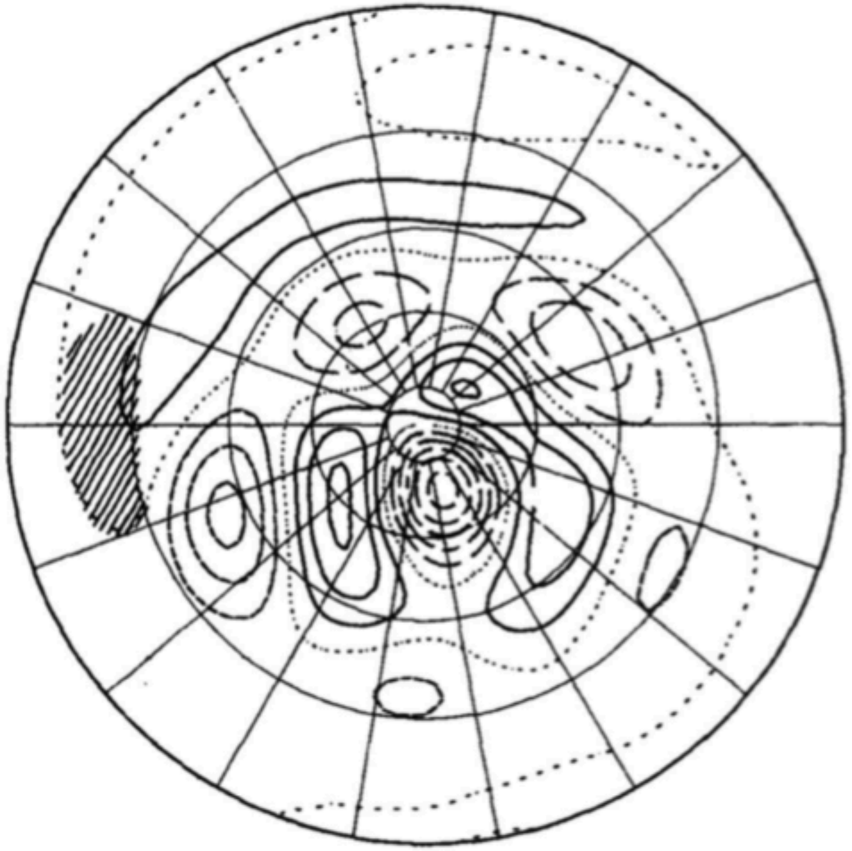
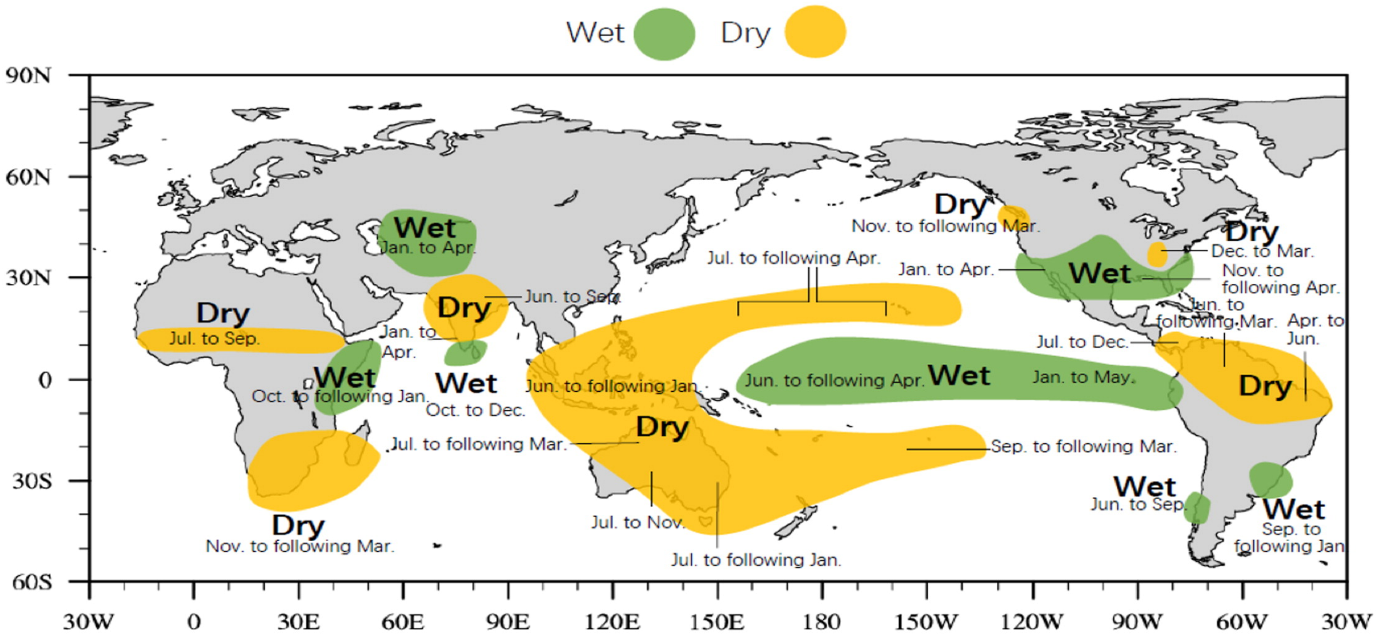


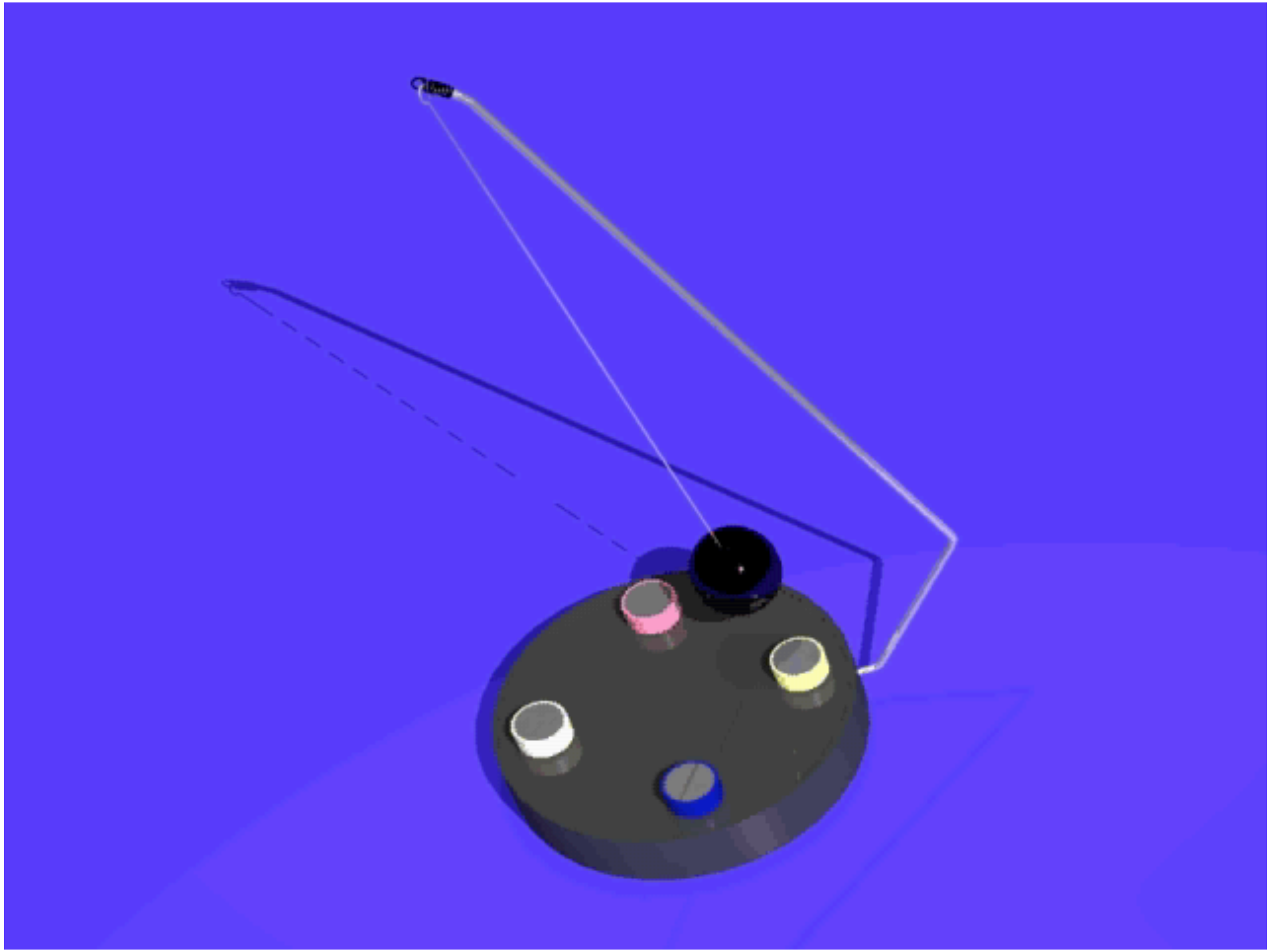
Figure 2. A 300 hPa geopotential height perturbation (contour interval 2 decameter) of a steady state linear solution of a five-layer primitive equation model forced by a deep elliptical heat source at 15°. The hatching represents the region of heating larger than 0.5 K/d. Adapted from *Hoskins and Karoly* [1981].

Global ENSO teleconnections

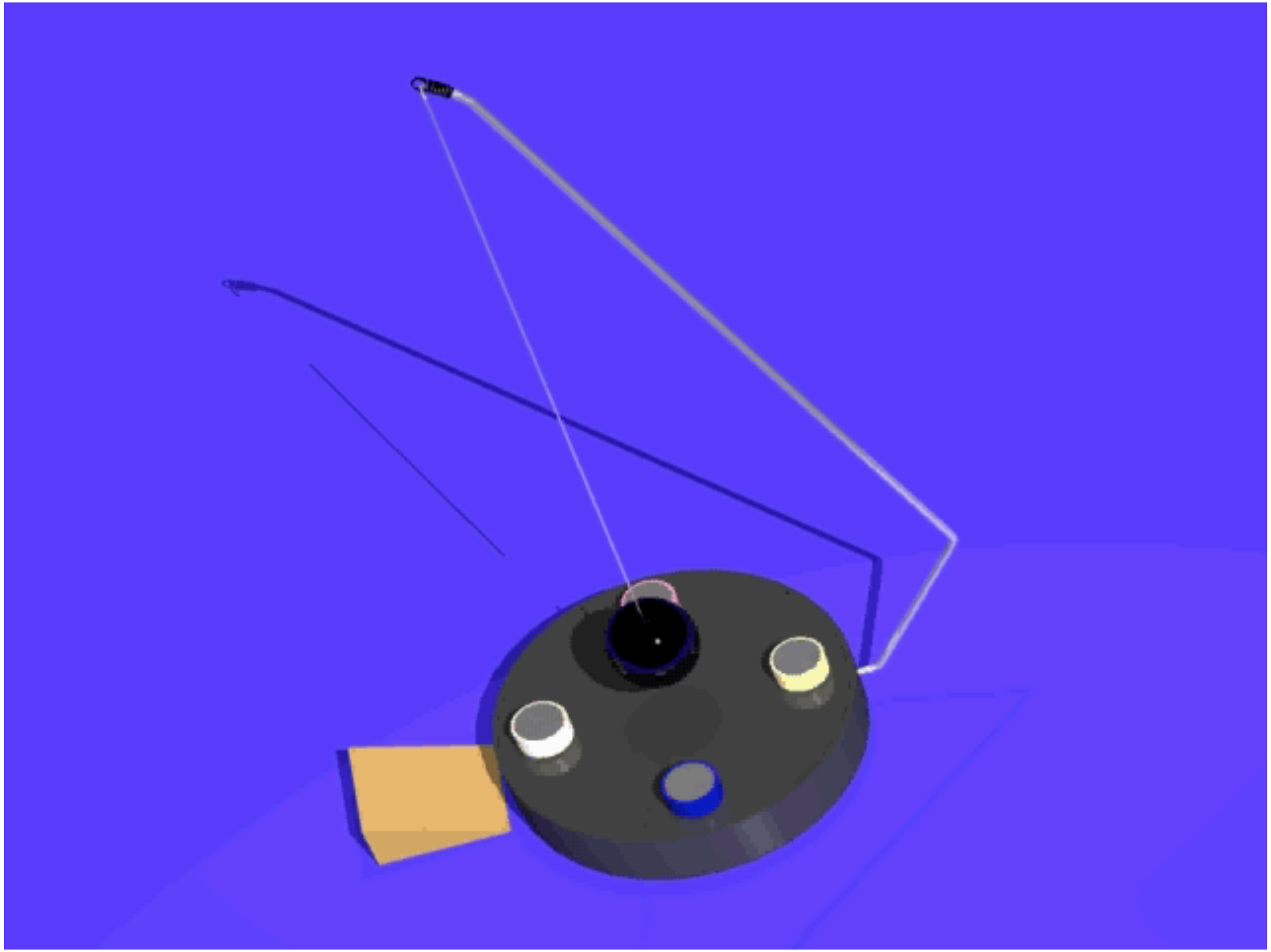


http://iri.columbia.edu/enso

Mechanical analogue of forcing and preferred circulation states

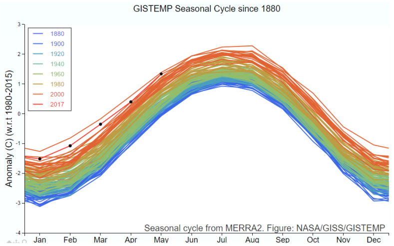


Mechanical analogue of forcing and preferred circulation states

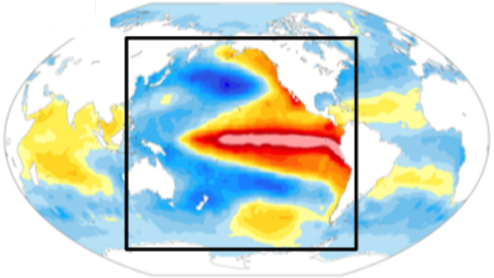


Sources of long-range predictability

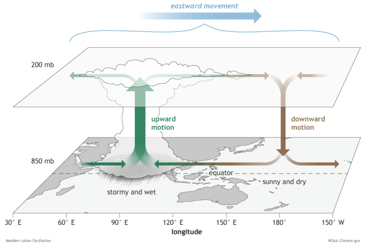
- **Seasonal cycle of solar radiation** (trivial?)



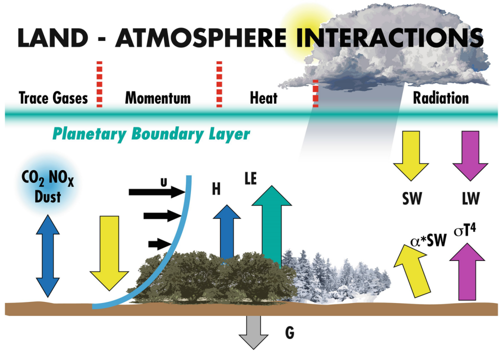
- **El Niño Southern Oscillation (ENSO)**



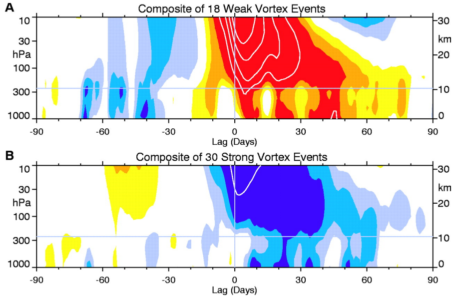
- **Other tropical ocean SSTs**



- **Local land surface conditions**

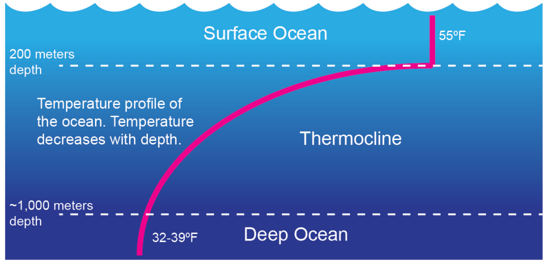


- **Stratospheric variability**



- **Sea-ice anomalies**

- **Mid-latitude ocean temperatures**



- **Volcanic eruptions**

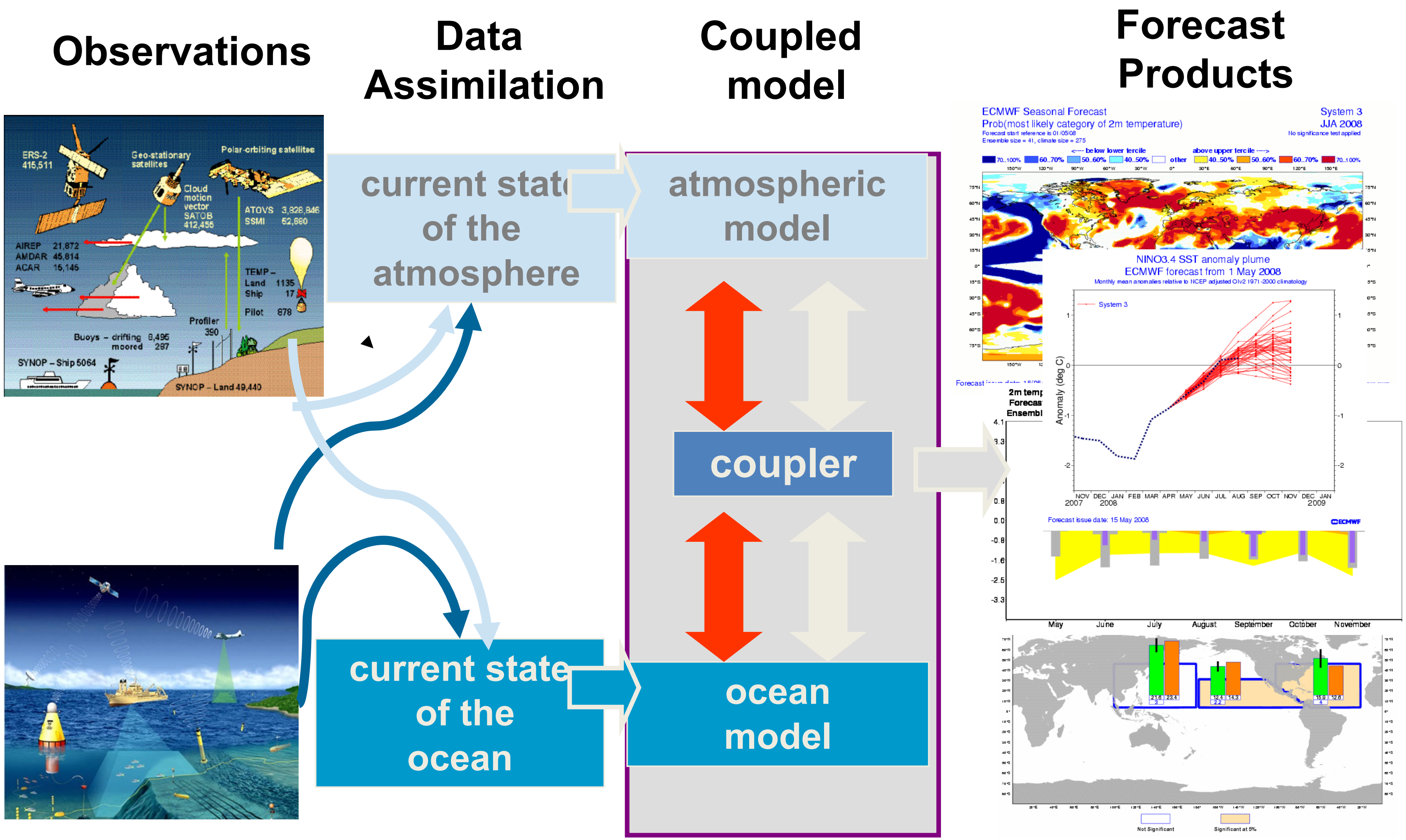
Regional anomalies in the atmosphere can persist for longer than the deterministic predictability limit

→ **substantial societal impacts**

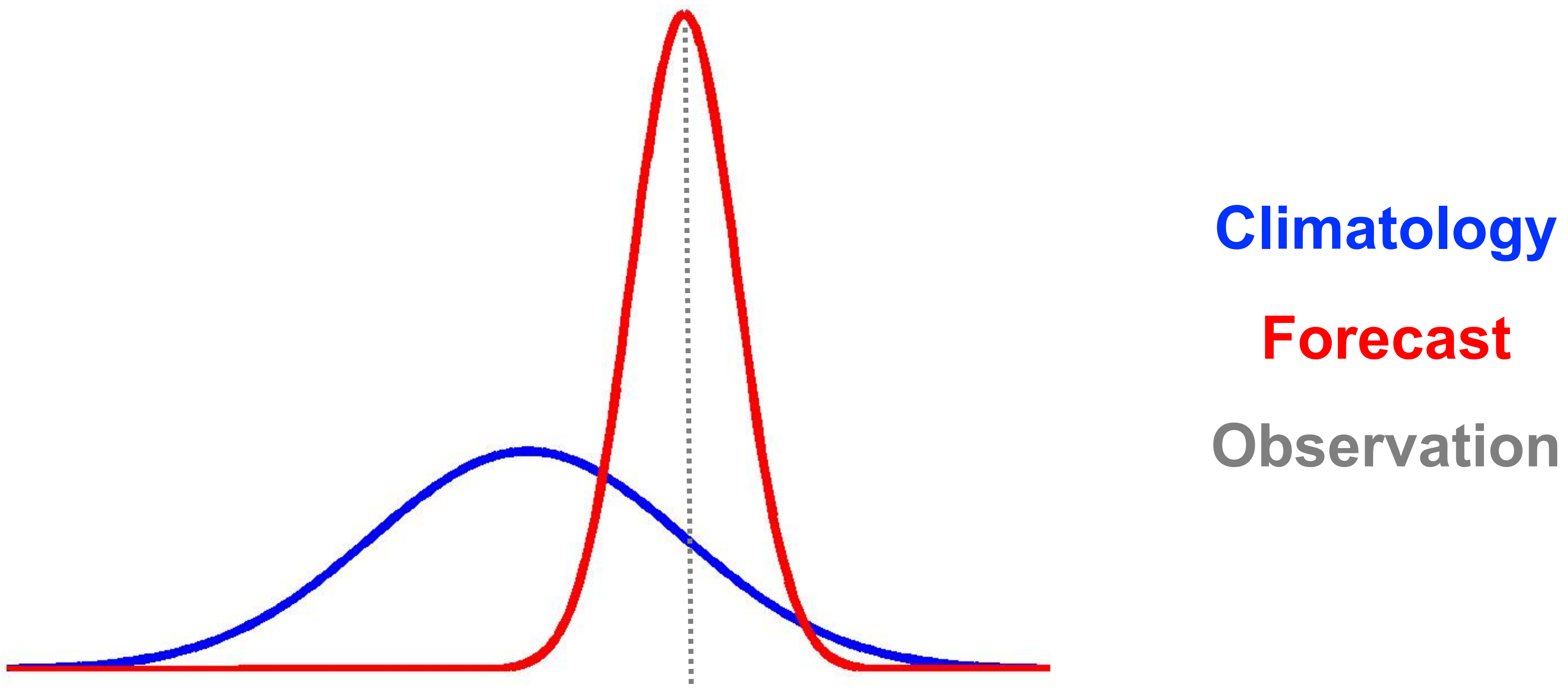
Prospect of predictability beyond that limit arises from interactions with slowly varying components of the climate system

→ **need to initialise and model coupled phenomena**

Forecast models for seasonal predictions

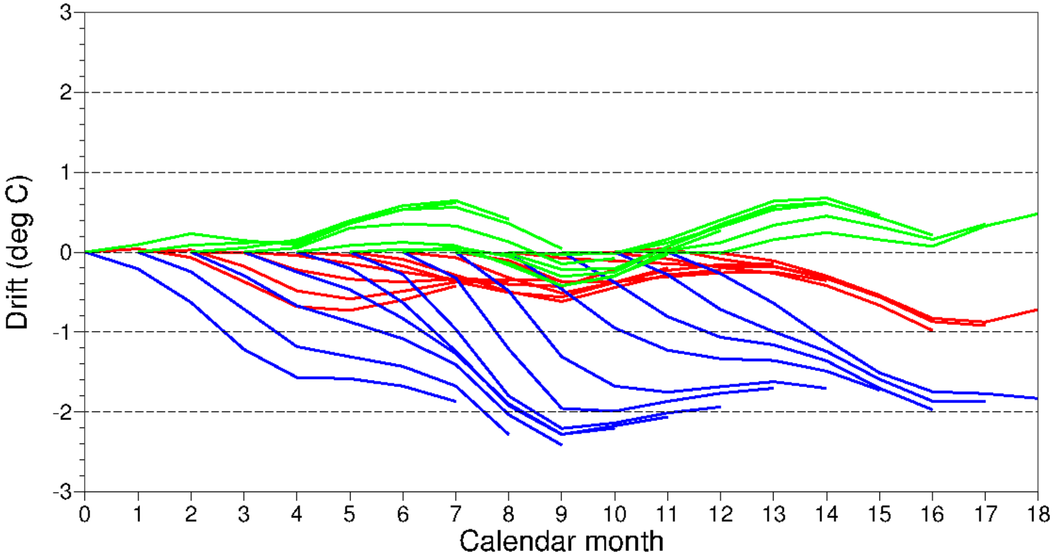


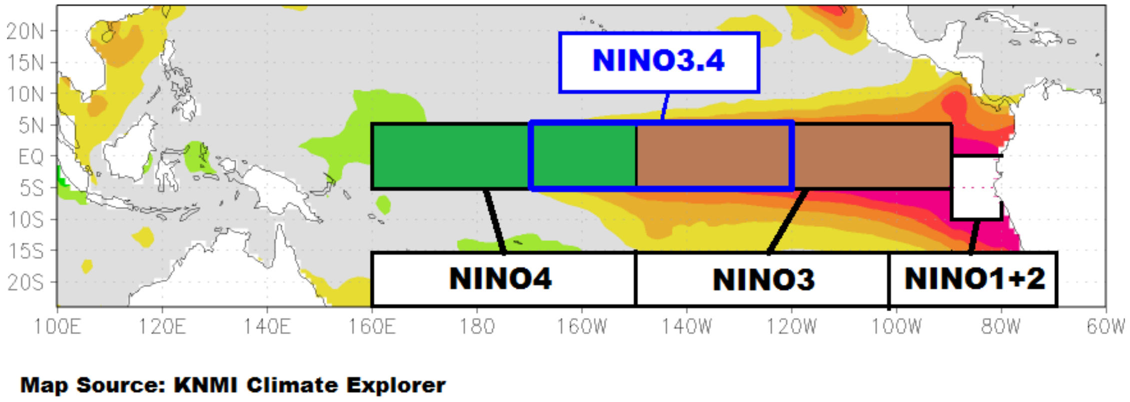
Aim: forecast distribution which sufficiently discriminates interannual signal from climatological background distribution

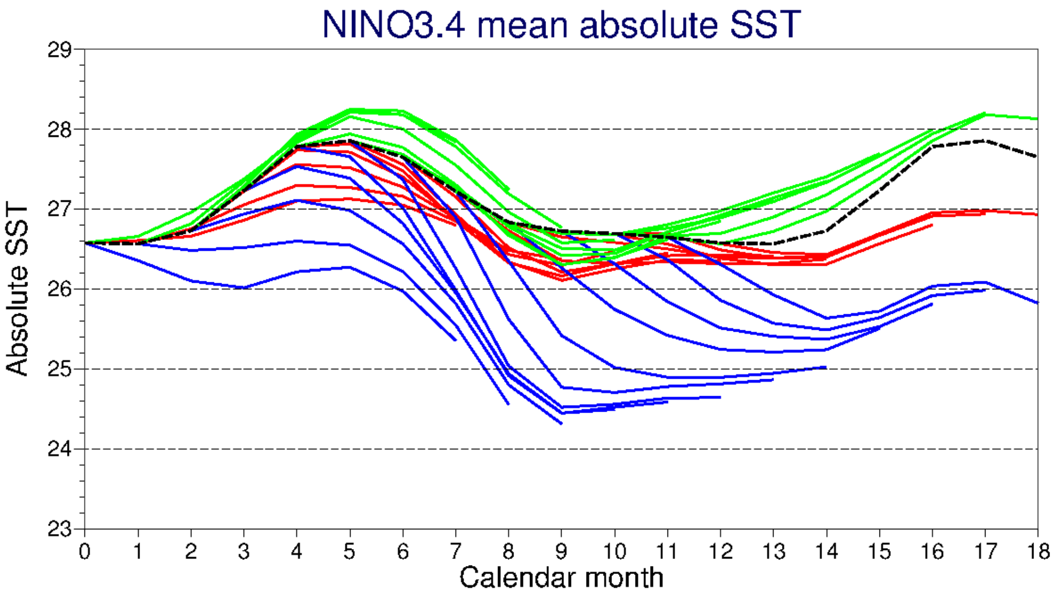


How well can we forecast ENSO?

Nini3.4 SST drift





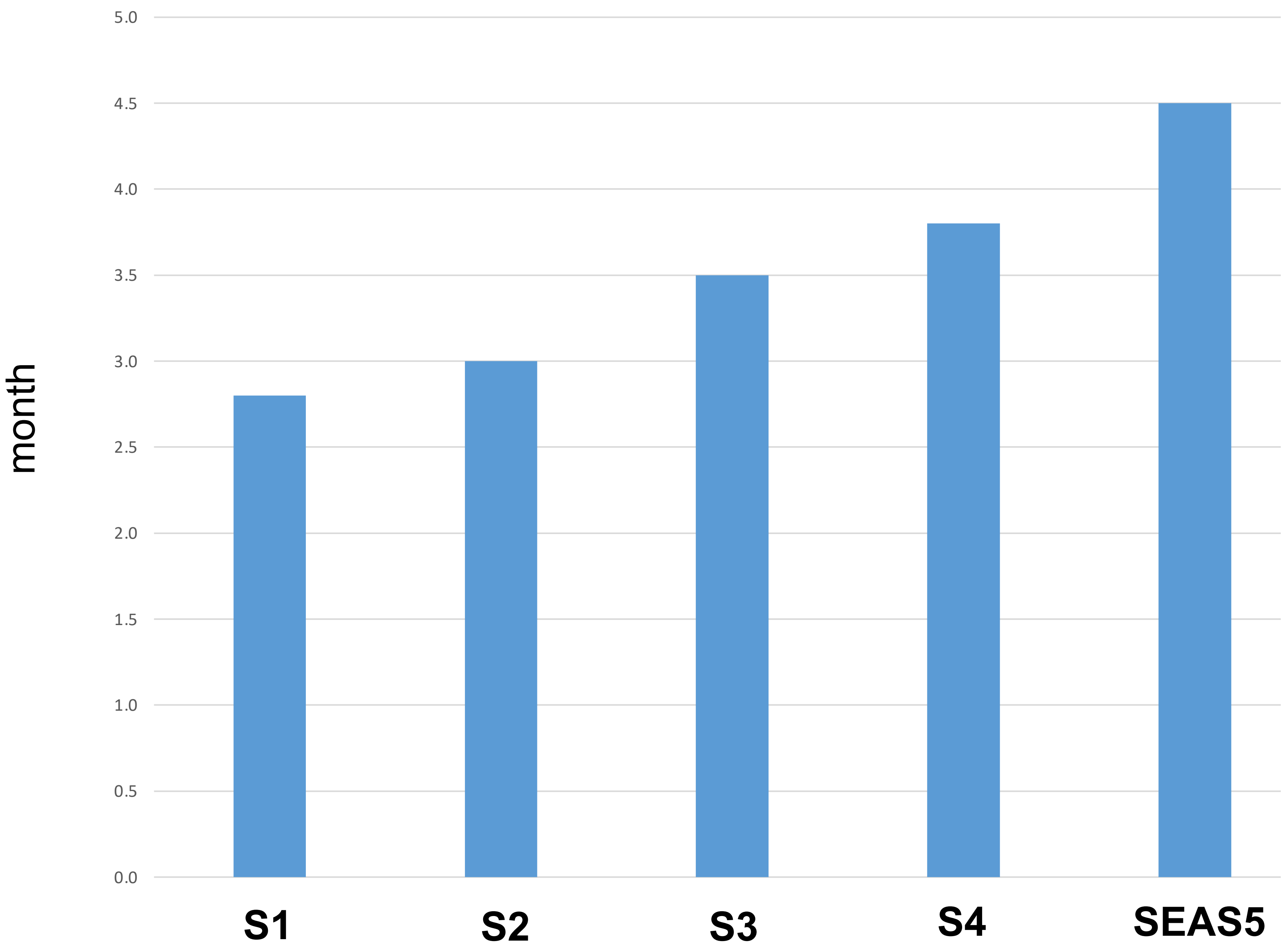


Local SST bias is a function of

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

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

Forecast lead time when Nino3.4 SST correlation skill < 0.9



Retrospective seasonal forecast over the period 1901-2010

1. ECMWF's atmospheric model with prescribed SSTs and sea-ice (ASF-20C)


Atmosphere


Land


Wave

Initial conditions from ERA-20C:
the ECMWF atmospheric reanalysis of the 20th Century

Weisheimer et al. (QJRMS 2017); O'Reilly et al. (GRL 2017); Weisheimer et al. (QJRMS 2019)

2. ECMWF's fully coupled atmosphere-ocean-sea-ice model (CSF-20C)


Atmosphere


Land


Wave


Ocean


Sea ice

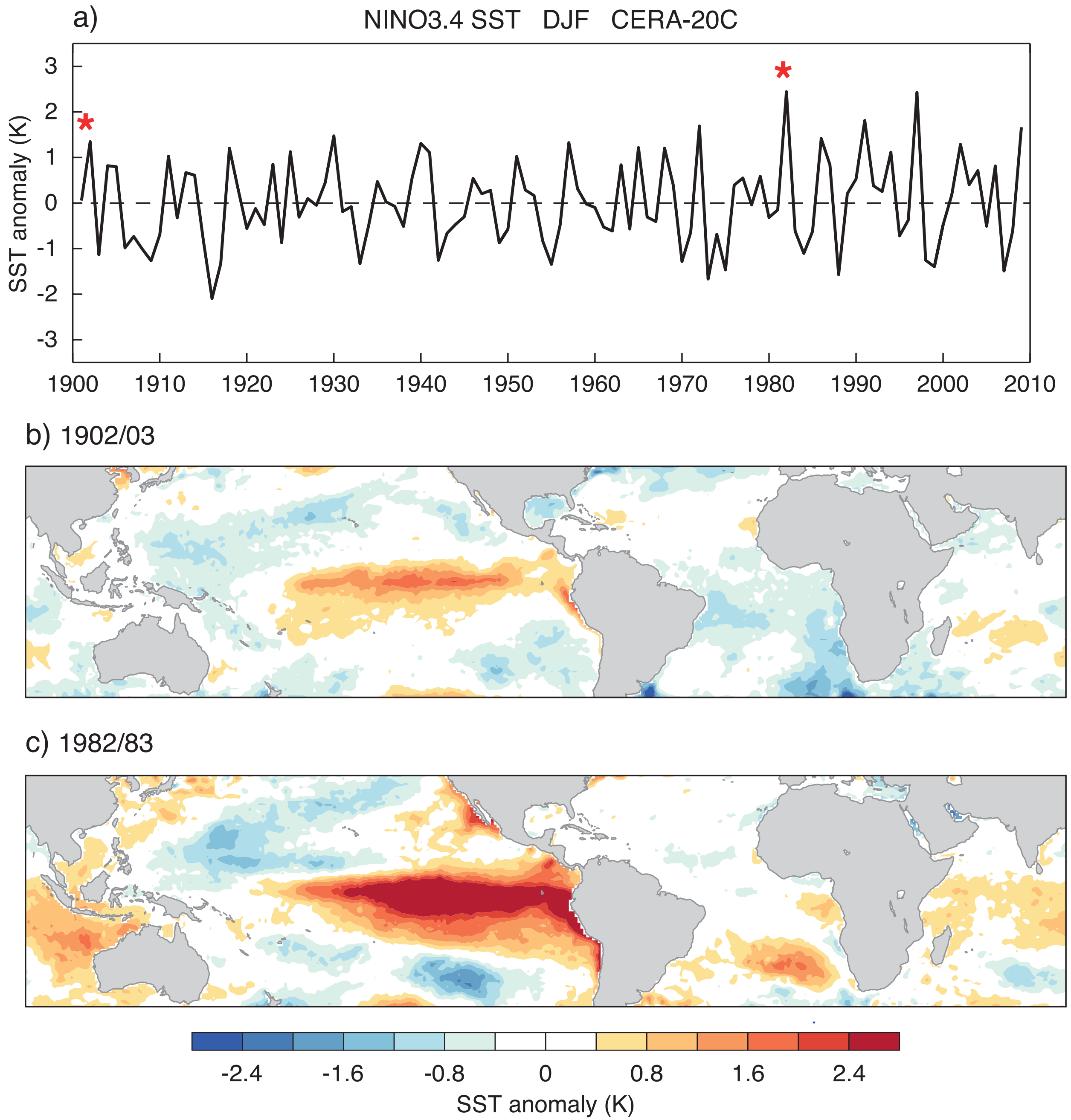
Initial conditions from CERA-20C: the first ECMWF coupled ensemble reanalysis of the 20th Century

Weisheimer et al. (BAMS 2020)

- IFS model cycle 41R1 (in-between S4 and SEAS5), T_L255L91 (ca. 60km)) + NEMO ORCA1L42 (1º) +LIM2
- Ensemble with 51 or 25 perturbed members
- 4-month forecast initialised on 1st of Feb/May/Aug/Nov each year (focus here: Nov → DJF)

Data are publicly available, see *Weisheimer et al. (BAMS 2020)*

ENSO during the 20th Century



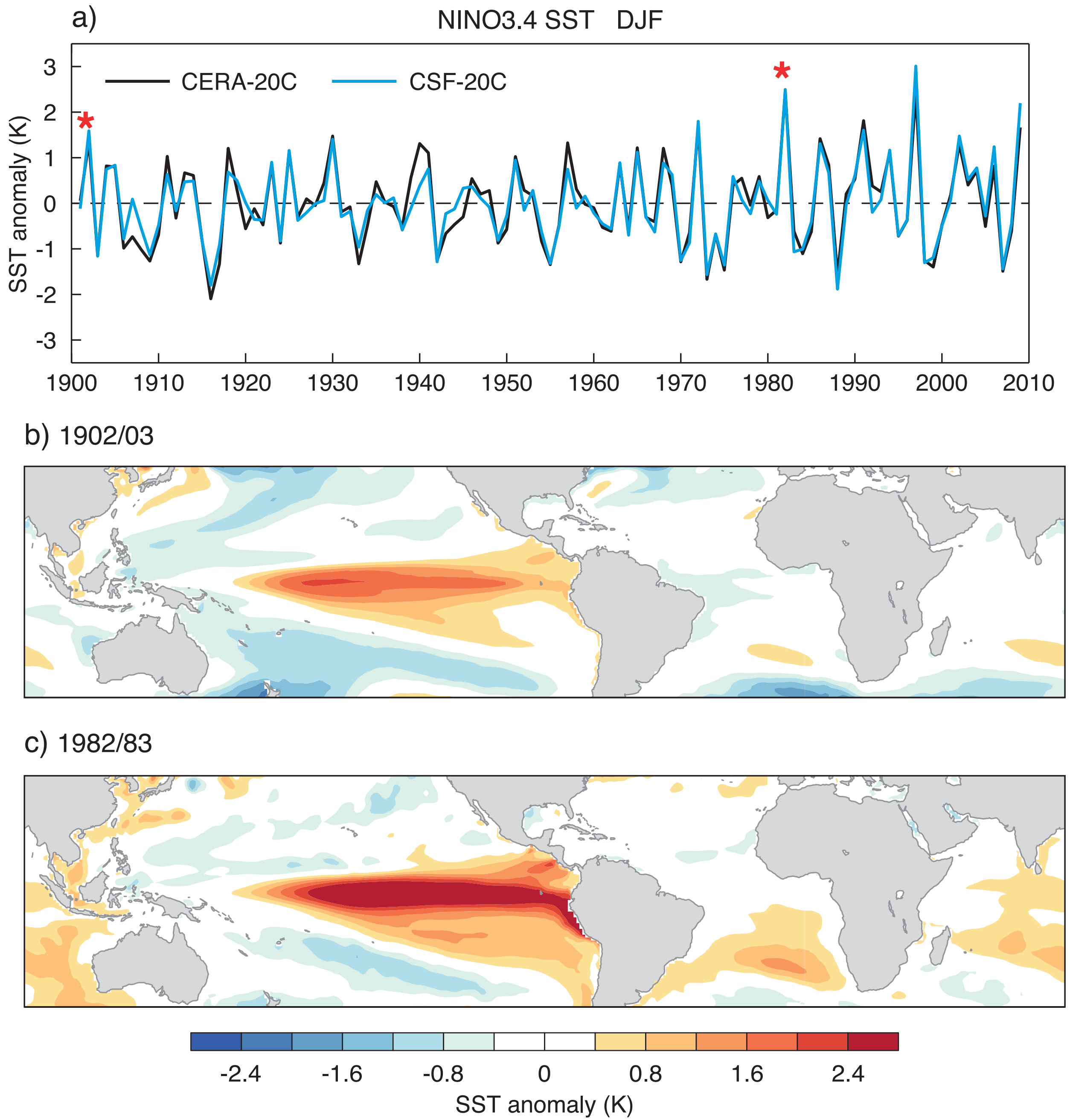
a) Time series of the NINO3.4 SST index anomaly in the central tropical Pacific during DJF in the CERA-20C reanalysis. Two examples of strong El Niño events near the beginning (1902/03) and the end (1982/83) of the 20th Century are depicted by the red stars.

b) The spatial structure of SST anomalies during the 1902/03 El Niño event in the CERA-20C reanalysis. The strong and large-scale warming of the central and eastern tropical Pacific is clearly visible. Cold anomalies develop in the western parts of the tropical Pacific.

c) As in b) but for the 1982/83 event

Weisheimer et al. (BAMS 2020)

ENSO in CSF-20C



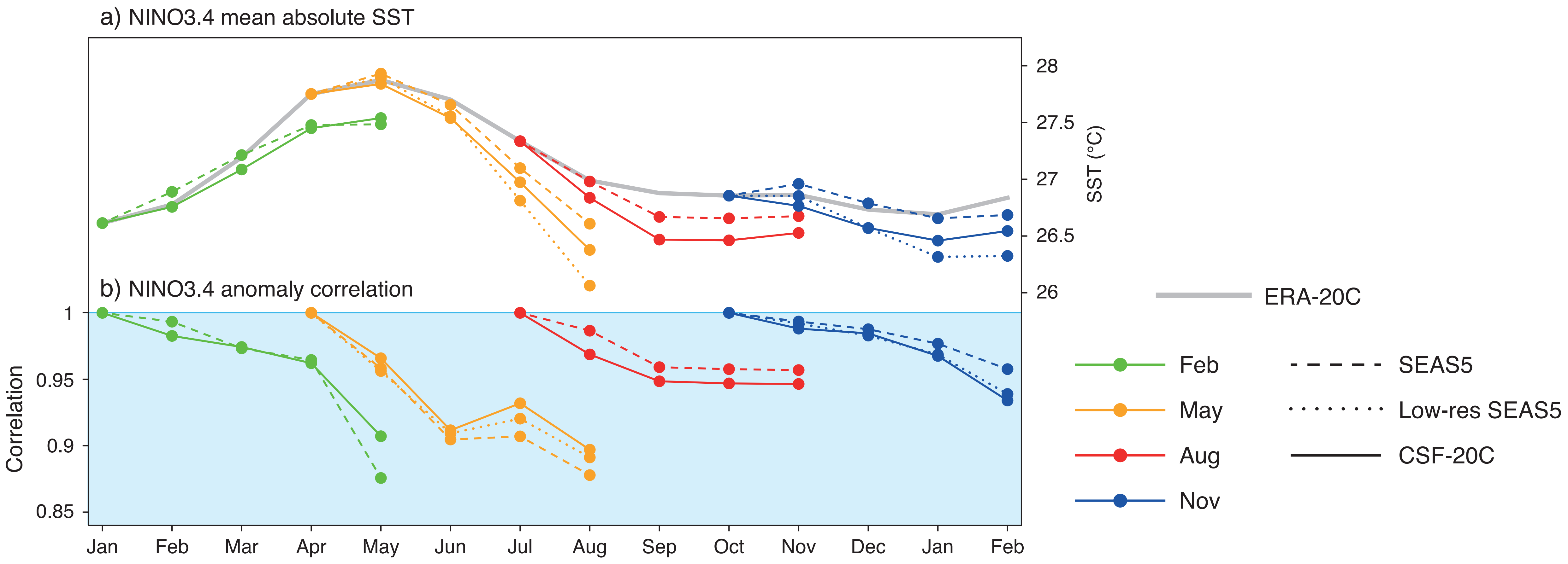
a) Time series of the NINO3.4 SST index anomaly in DJF in the CERA-20C reanalysis (black) and ensemble mean of the CSF-20C hindcasts (blue).

b) The spatial structure of SST anomalies during the 1902/03 El Niño event in the CSF-20C hindcasts. There is a very good agreement with the strong and large-scale warming of the central and eastern tropical Pacific in CERA-20C. The cold anomalies in the western parts of the tropical Pacific are also visible.

c) As in b) for the for 1982/83 event

Weisheimer et al. (BAMS 2020)

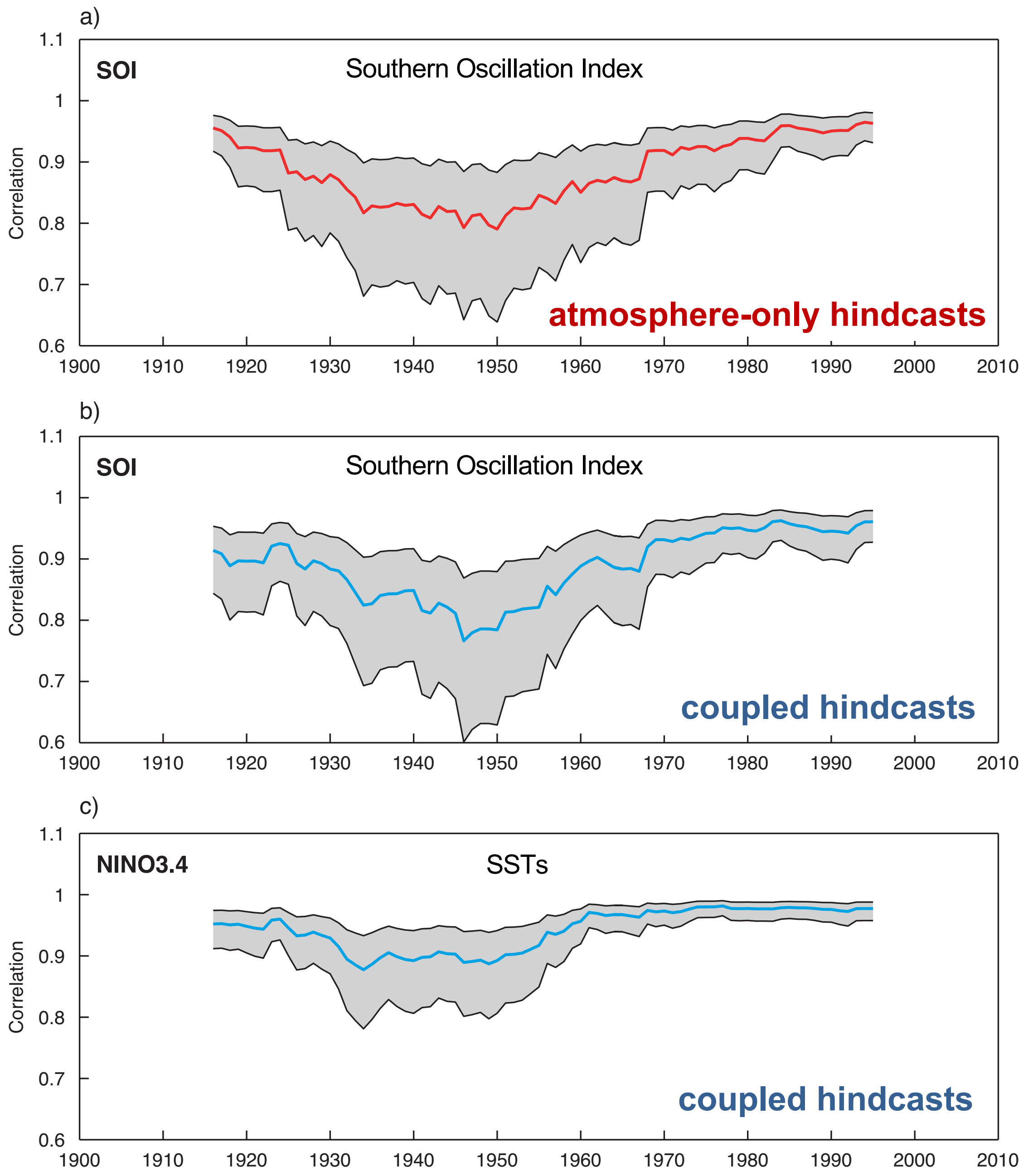
ENSO forecasts: drift and skill



Drift (upper part) and anomaly correlation skill (lower part) of predictions of the NINO3.4 SST index for different start dates throughout the calendar year during the hindcast period 1981-2009. CSF-20C data are shown in solid colored lines, SEAS5 in dashed lines and the lower-resolution SEAS5 in dotted lines (only available for May and November start dates). Different colors indicate different start dates of the hindcasts. The gray curve in the upper part shows the climatological mean evolution of the SSTs in ERA-20C over this period.

Weisheimer et al. (BAMS 2020)

ENSO forecast skill during the 20th Century



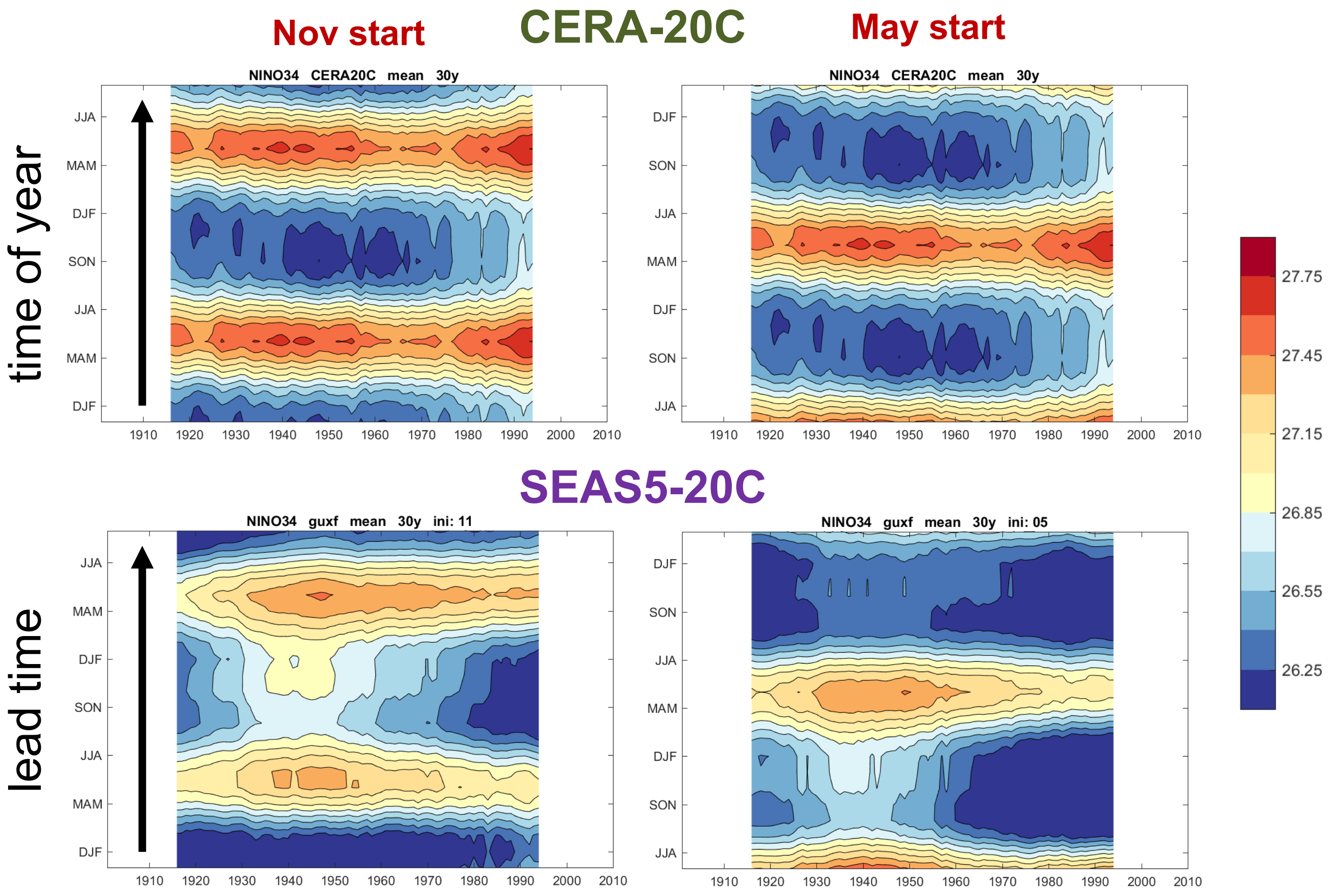
Decadal variability of ENSO forecast skill

- DJF from forecasts initialised in November
- 30-year moving window correlation coefficients of the hindcast ensemble mean with reanalysis (CERA-20C), plotted at central year
- gray shaded bands indicate the 5- 95% confidence intervals

Weisheimer et al. (BAMS 2020)

Motivation

- Can we predict ENSO beyond one year? Has the model climate converged after 2 years?
- How flow-dependent is the predictability of ENSO on seasonal to multi-annual timescales in the presence of multi-decadal climate variability?

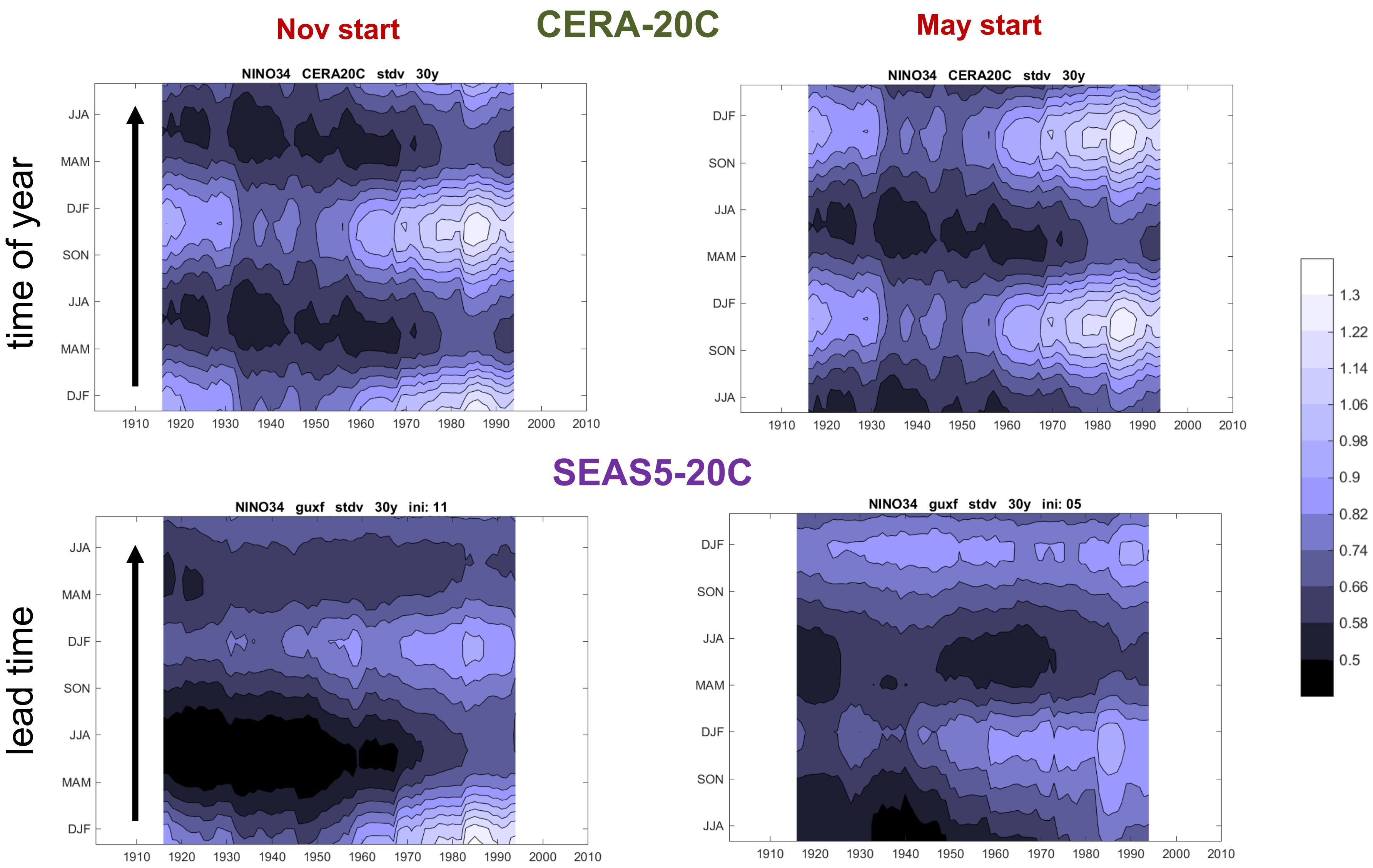


30-year moving averages of NINO3.4 SST mean state across the 20th Century

Experiments

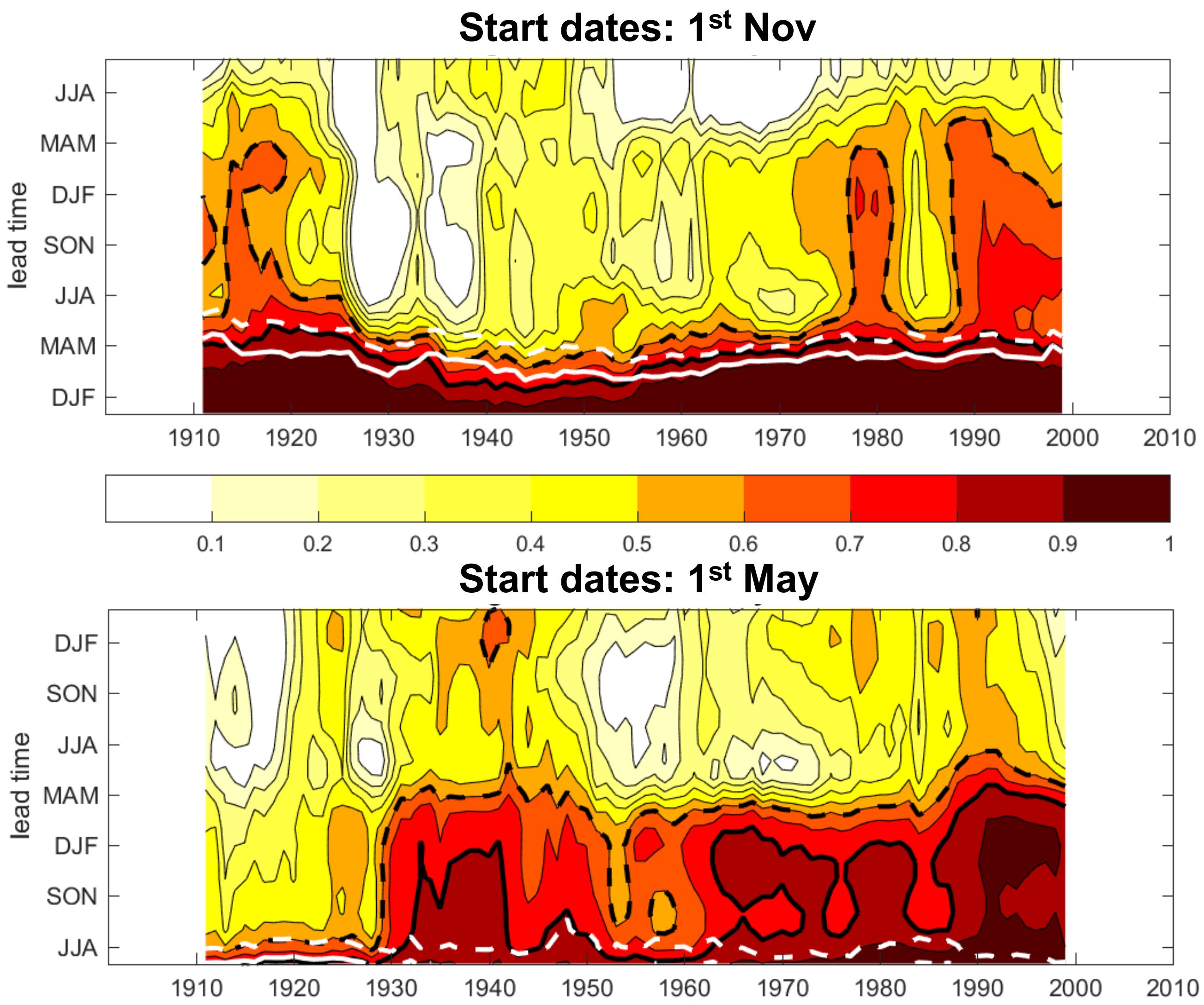
- Coupled hindcasts initialised from coupled 20th Century reanalysis CERA-20C from 1901 to 2010
- SEAS5 low-res model resolution: T_{co}199L91 (ca. 50km) with ORCA1Z42 (1 degree)
- 24-month forecasts with 10 ensemble members
- additional experiments to test sensitivity to ocean initial conditions (impact of data assimilation and and atmospheric forcing)

SEAS5-20C NINO3.4 SSTs: variability



30-year moving averages of NINO3.4 SST standard deviation (amplitude) across the 20th Century

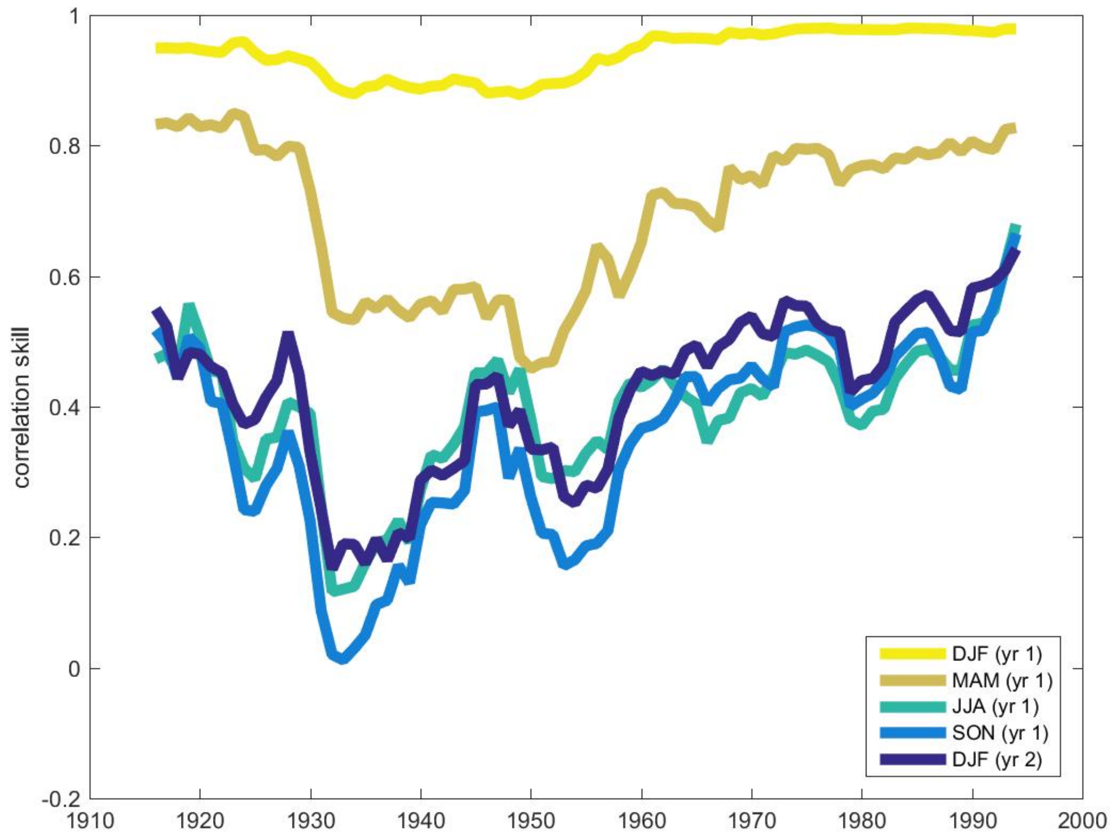
SEAS5-20C ENSO: skill evolution



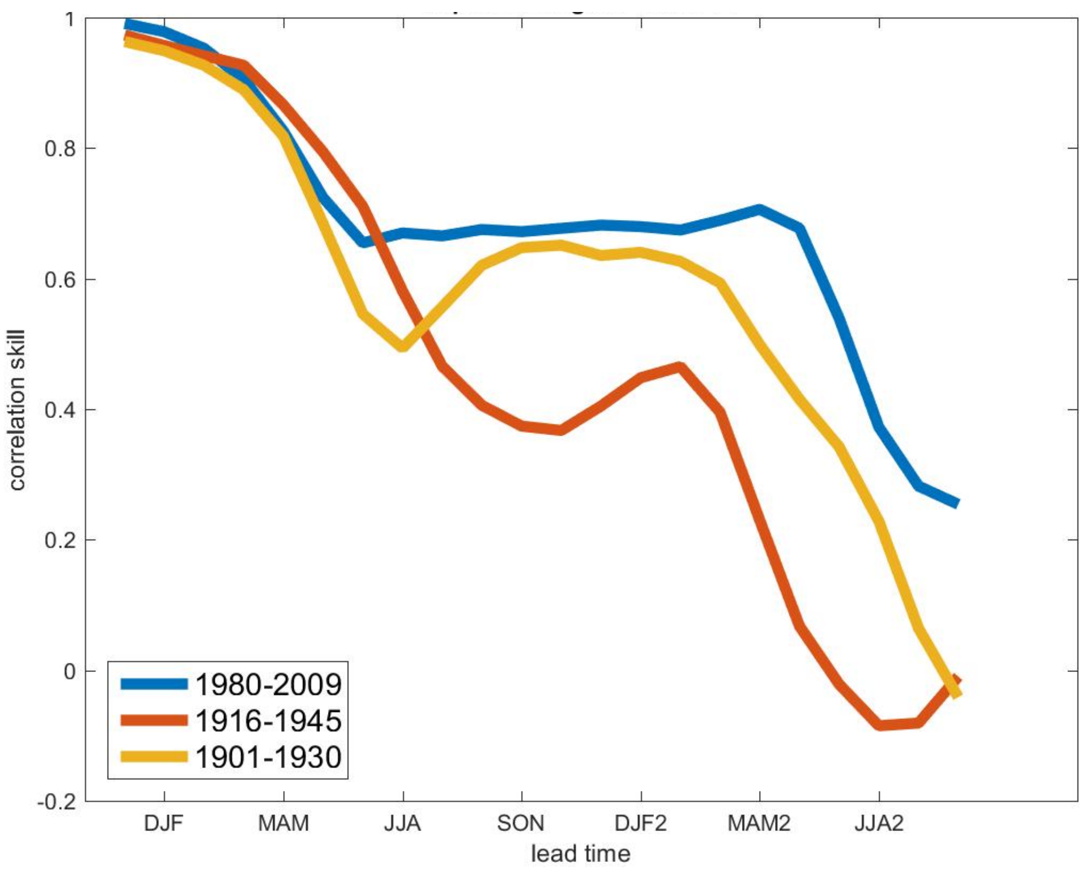
NINO3.4 SST 20-yr moving correlation skill vs. CERA-20C as a function of hindcast epoch and lead time (y-axis, lead time moves up). The black and white contour lines indicate skill of 0.8 (solid) and 0.6 (dashed) for the experiment (black) and persistence (white).

SEAS5-20C NINO3.4 SSTs: skill evolution

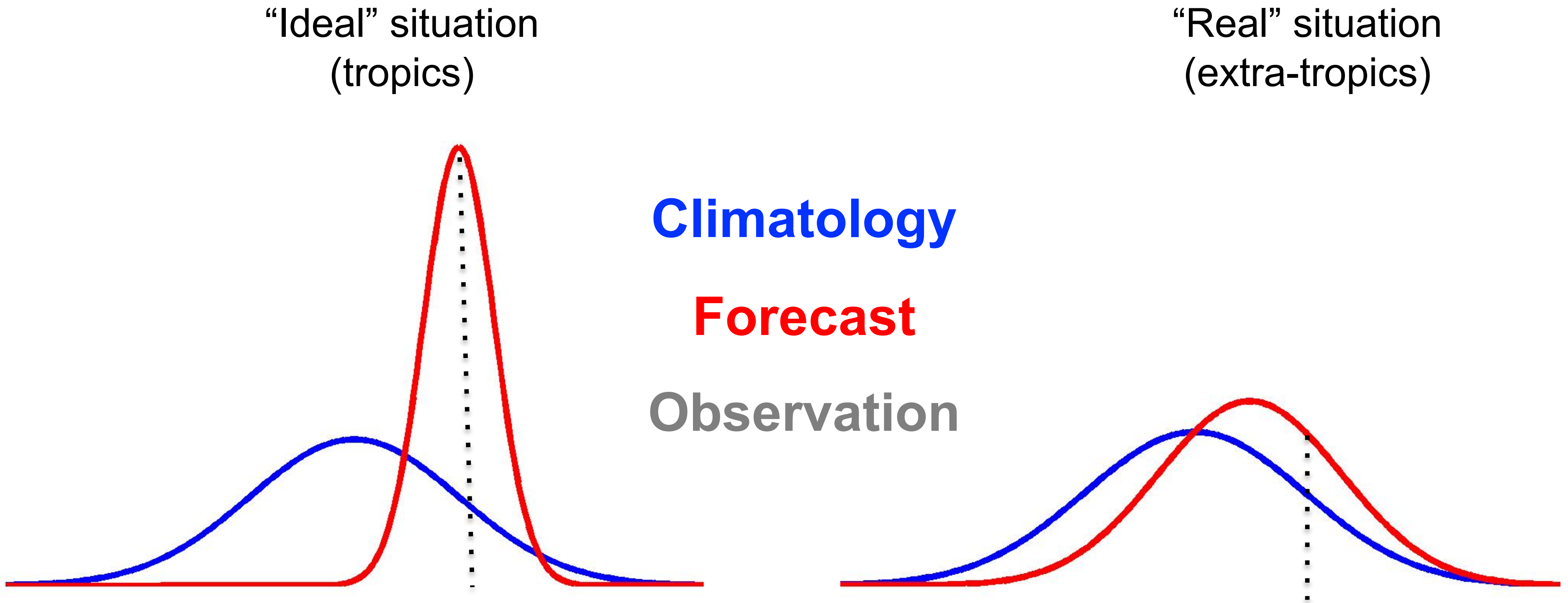
Correlation skill as a function of hindcast epoch for different lead times



Correlation skill as a function of lead time for different hindcast epochs



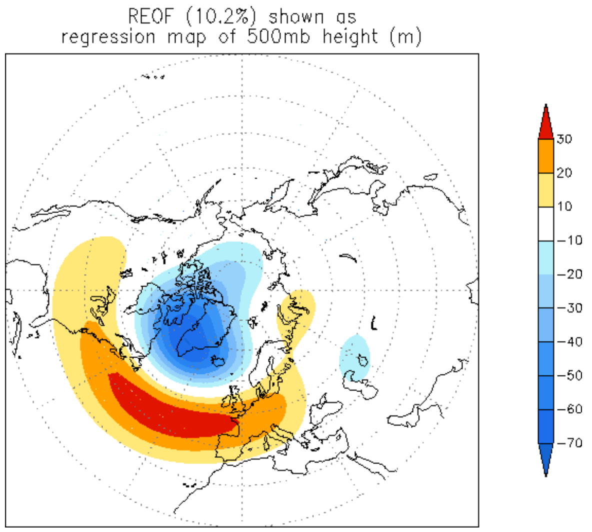
Aim: forecast distribution which sufficiently discriminates interannual signal from climatological background distribution

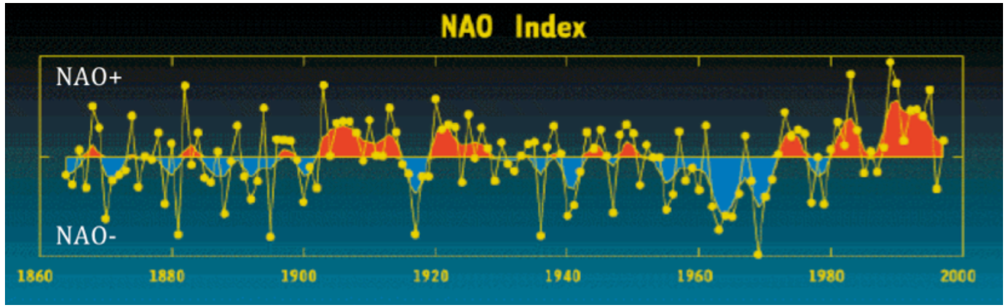


Extratropical modes of variability

North Atlantic Oscillation (NAO)

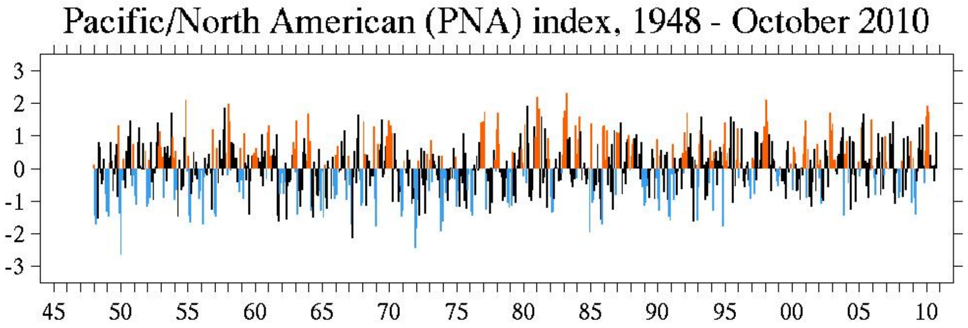
- Dominant mode of variability on a range of time scales over the North Atlantic-European region
- Typically defined as the 1st EOF of MSLP or Z500
- NAO index: 1st Principal Component or sometimes (mostly for historical reasons) as normalised MSLP difference between Iceland and the Azores

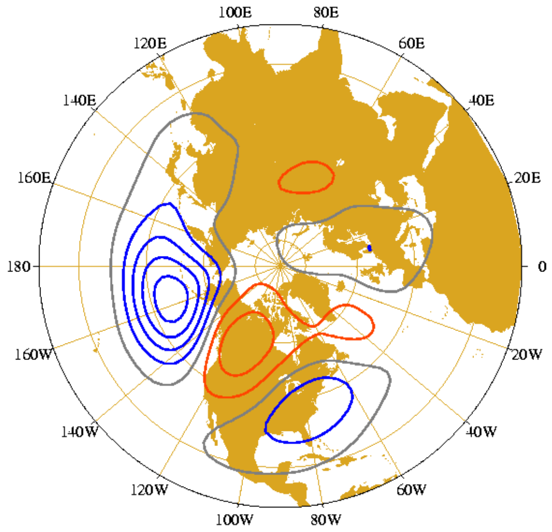




Pacific North America pattern (PNA)

- Prominent natural mode of low-frequency variability in the NH extratropics
- Strongly influenced by ENSO





Wallace & Gutzler (MWR 1981)

Seasonal forecasts of the weather and climate over the Euro-Atlantic region are difficult due to

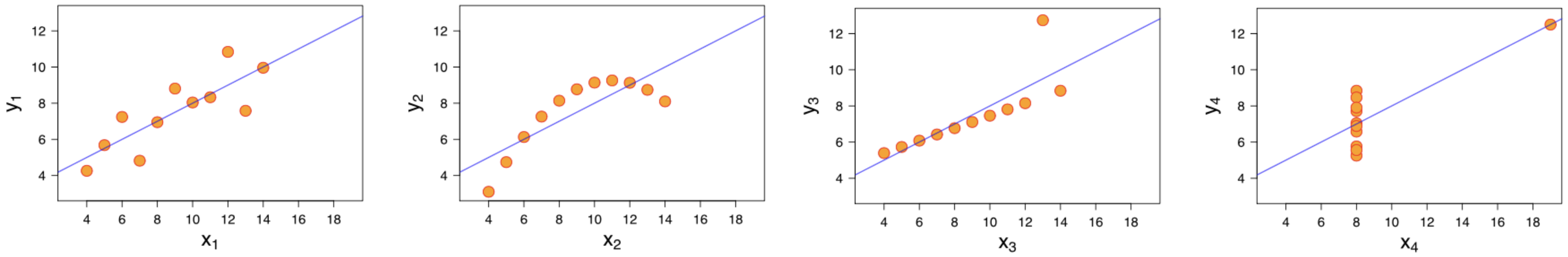
- low signal-to-noise ratios in predictability of extratropical atmosphere
- teleconnections from tropical forcings are weaker, and perhaps more manifold, than for other areas in the world
- sample sizes are intrinsically small (mainly limited by number of observed seasons, usually $\mathcal{O}(30)$)

Estimates of seasonal predictability, skill and reliability suffer from rather large uncertainties

Illustration of correlation pitfalls after Anscombe (1973):

- Four pairs of *x*-*y* variables with all *y* variables having the same mean (=7.5) and variance (=4.1)
- Sample size is 11 for each and the correlation between *x* and *y* is 0.82 in all four samples
- However, the distributions of variables are very different

Anscombe's quartet



Normally distributed, "well behaved"

Not normally distributed, non-linear relationship

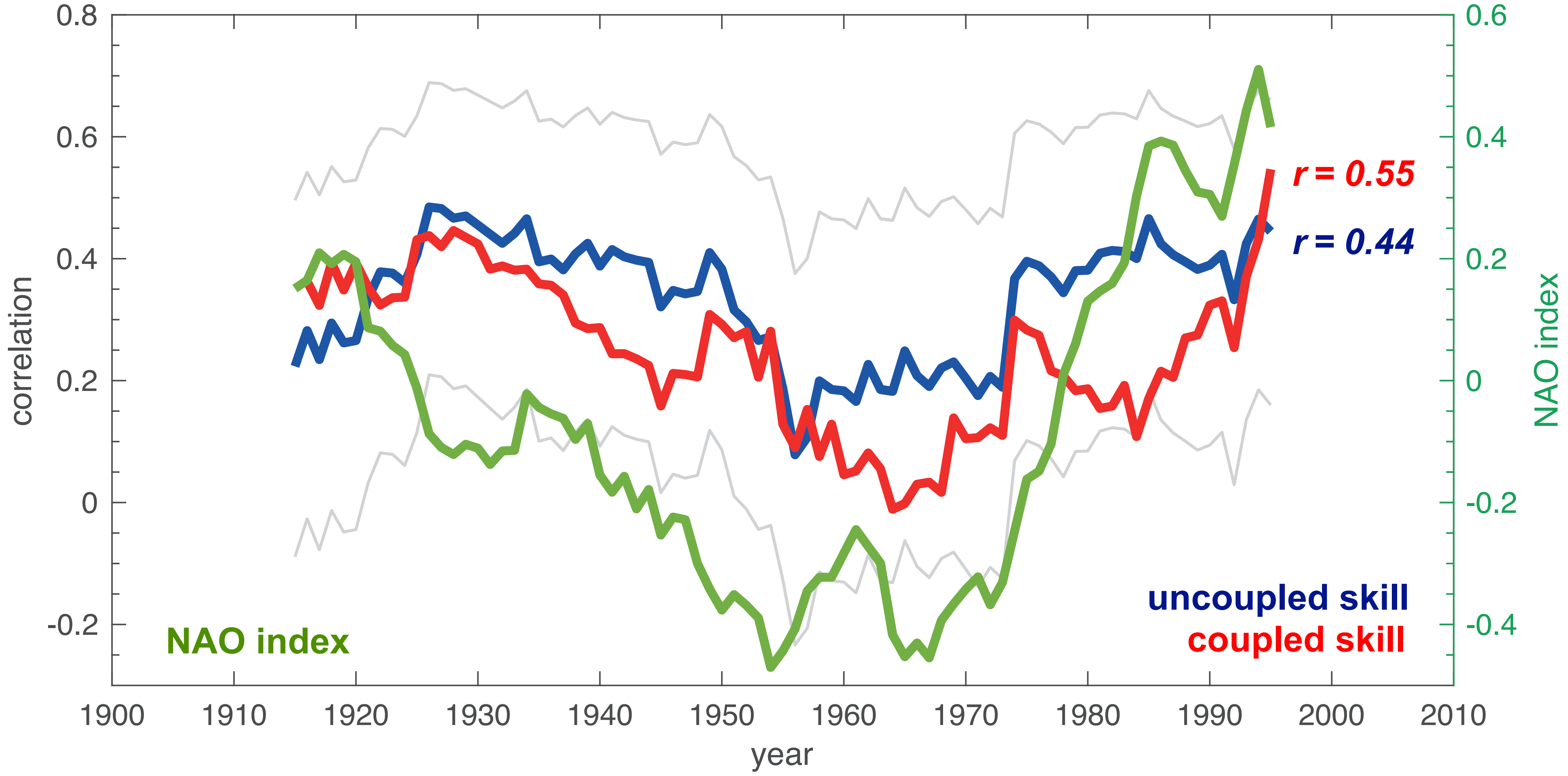
Perfect linear relationship except for one outlier

No relationship with one outlier

Anscombe (Amer. Statist. 1973)

Predictability of the NAO during the 20th Century

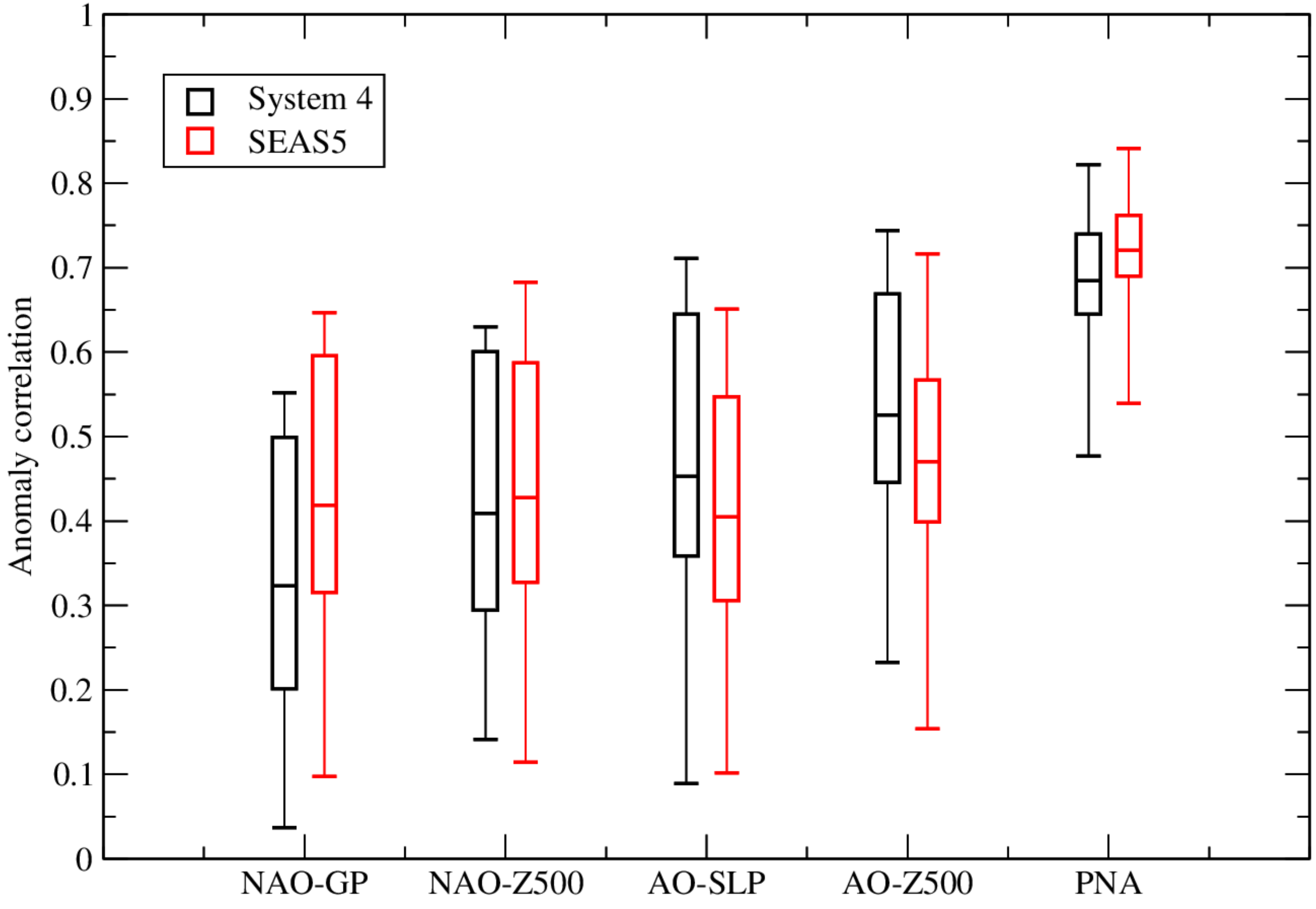
$r_{1901\text{-}2009} = 0.31$ $r_{1901\text{-}2009} = 0.28$



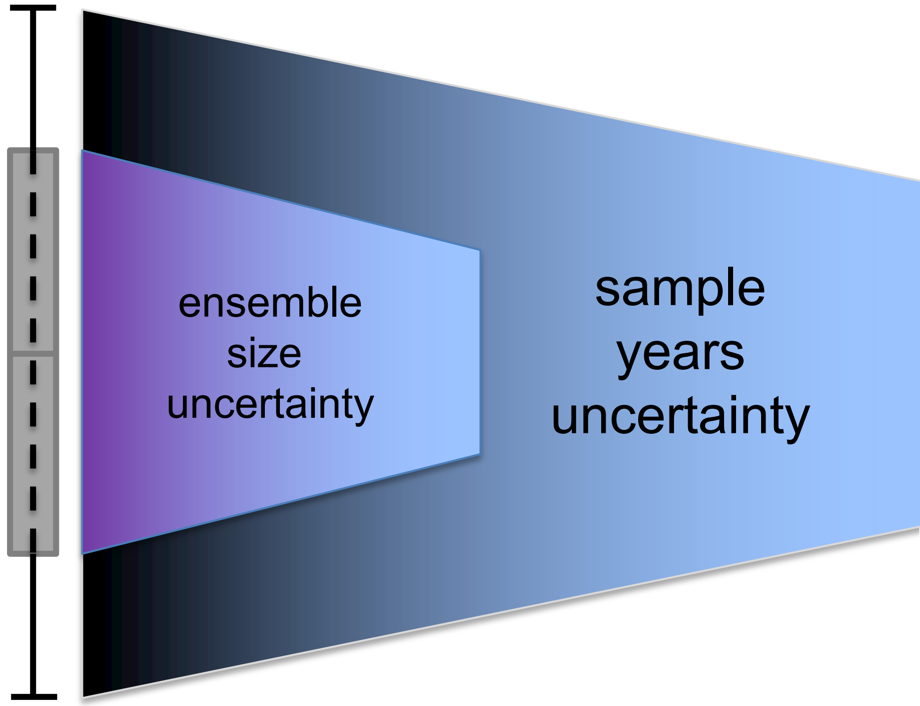
Sample uncertainty and correlation skill

Hindcast skill of the NAO, AO and PNA in DJF

Nov start dates 1981-2016, 51 ensemble members

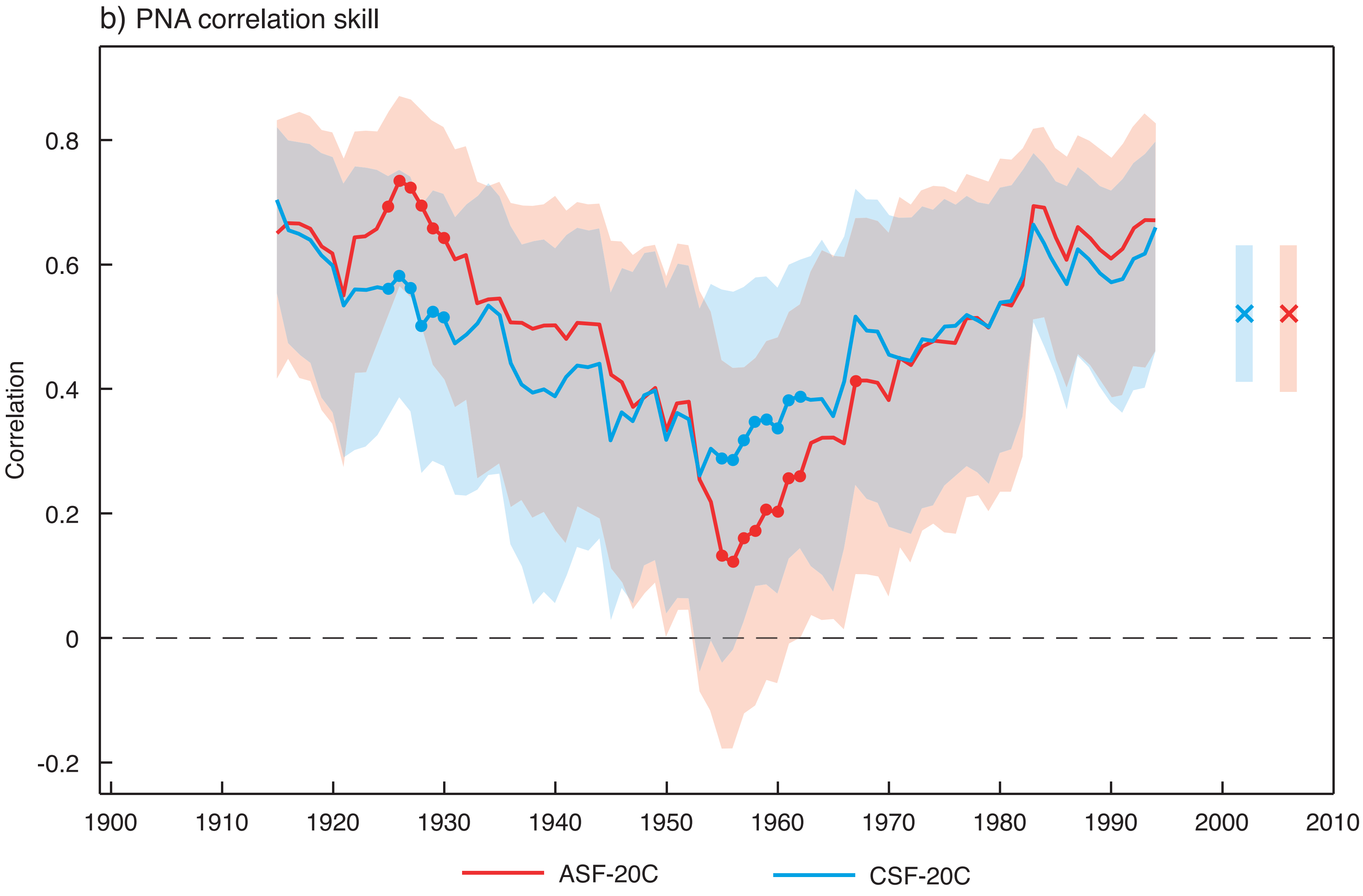


Courtesy Tim Stockdale



Two types of sampling uncertainty

Multidecadal variability of PNA forecast skill

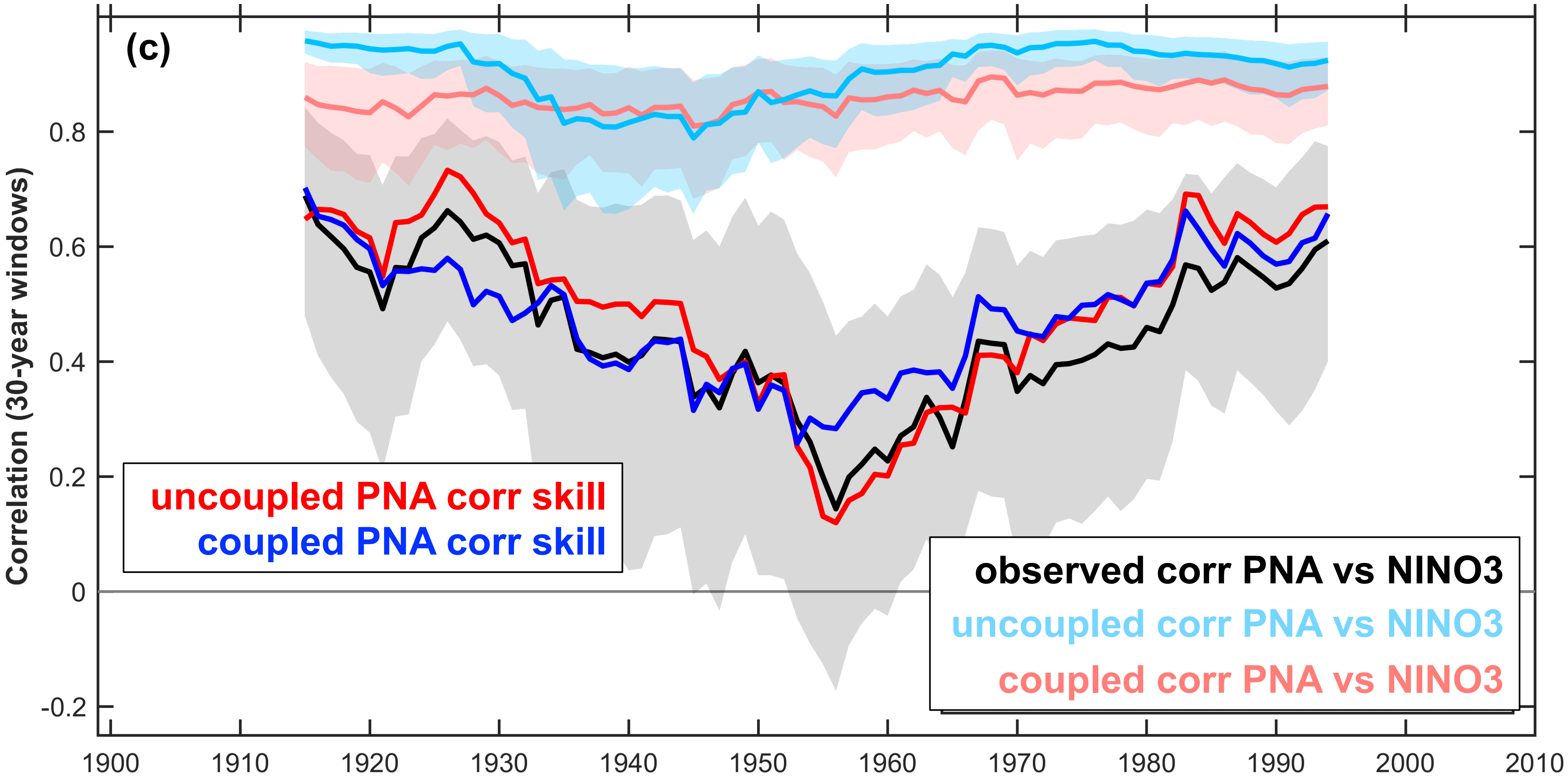


Weisheimer et al. (BAMS 2020)

Multidecadal variability of ENSO - PNA teleconnections

- Weakening of the obs. relationship between NINO3 SSTs and PNA during the mid-Century period
- Ensemble mean model PNA response to NINO3 SSTs is very stable over time (no weakening)

→ Lack of PNA skill in the mid-Century period?



O'Reilly et al. (GRL 2017)
Weisheimer et al. (BAMS 2020)

Signal and noise

RPC – Ratio of Predictable Components
(the signal-to-noise "paradox")

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

see *Eade et al (GRL, 2014)*

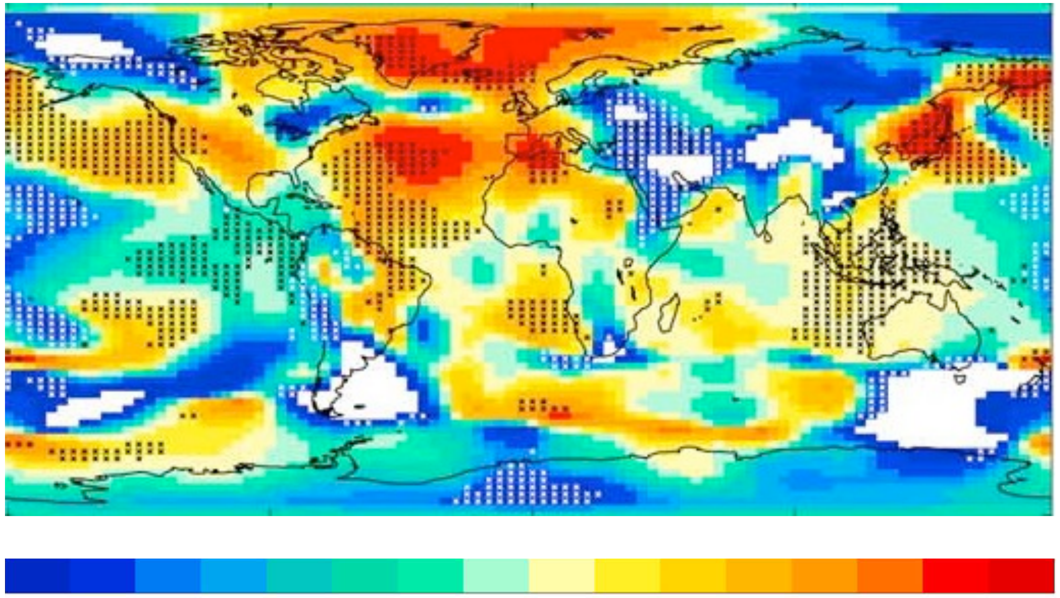
PC: predictable parts of the total variance

PC_{obs} : predictable component of the observations

PC_{model}: predictable component in the model

- Perfect model ensemble: **RPC==1**
- **RPC > 1** → underconfidence (overdispersive); model underestimates real-world predictability
- **RPC < 1** → overconfidence (underdispersive); model predictability is larger than in real world

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



Eade et al. (GRL 2014)

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

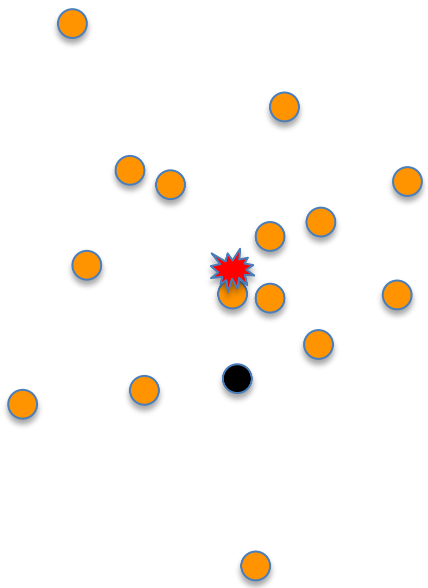
See also Scaife & Smith (npj Clim. Atmos. Sci. 2018), Smith et al. (Nature 2020)

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

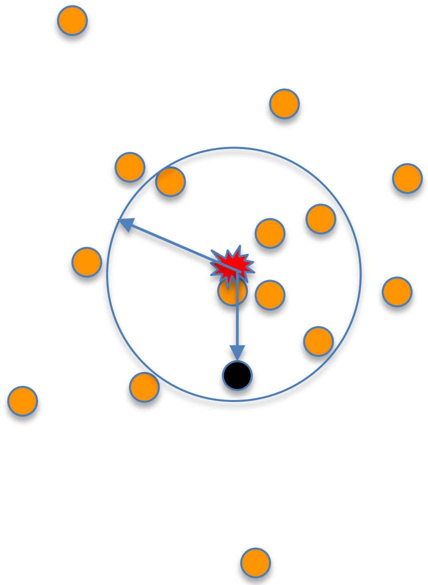
→ Replace observation with ensemble member



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*

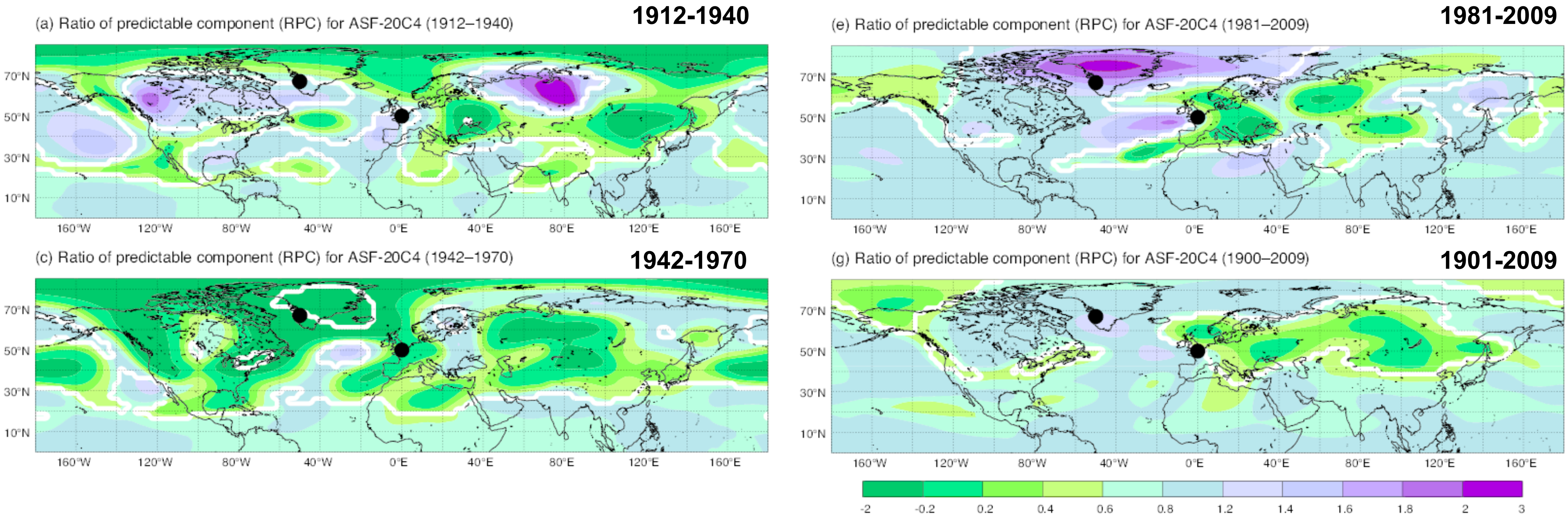
  ensemble spread > RMSE → ensemble is *overdispersive*

- 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

Ratio of Predictable Components (RPC) during different hindcast epochs

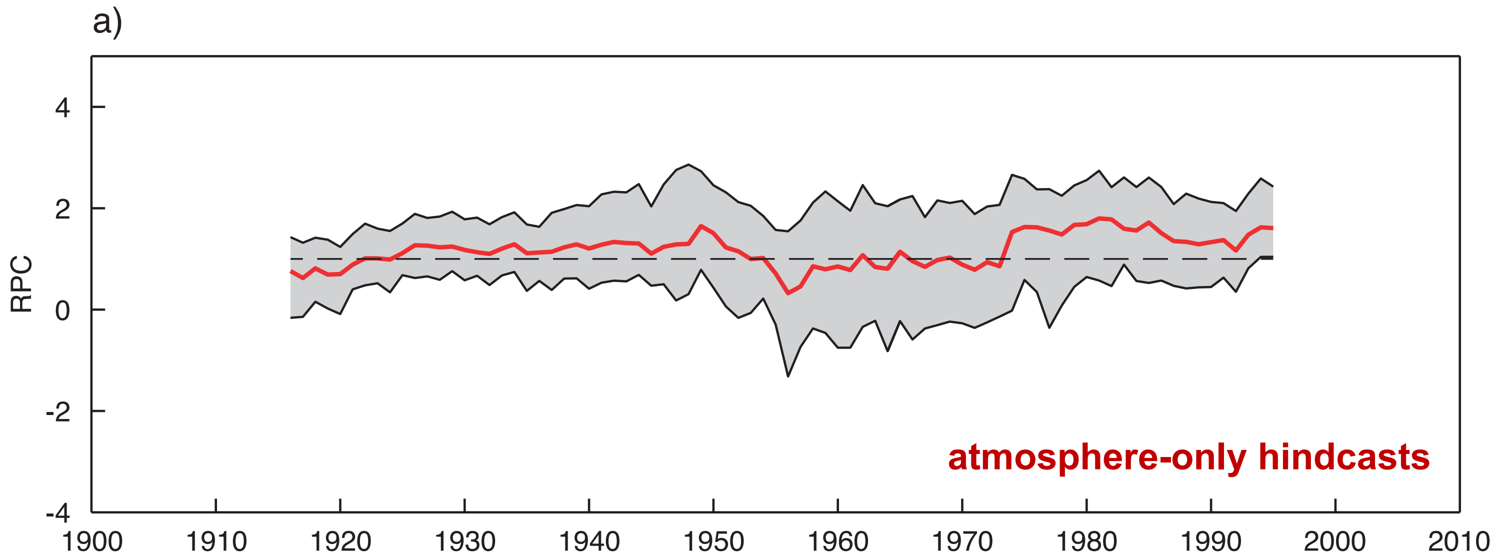


Weisheimer et al. (QJRMS 2019)

Signal and noise: NAO

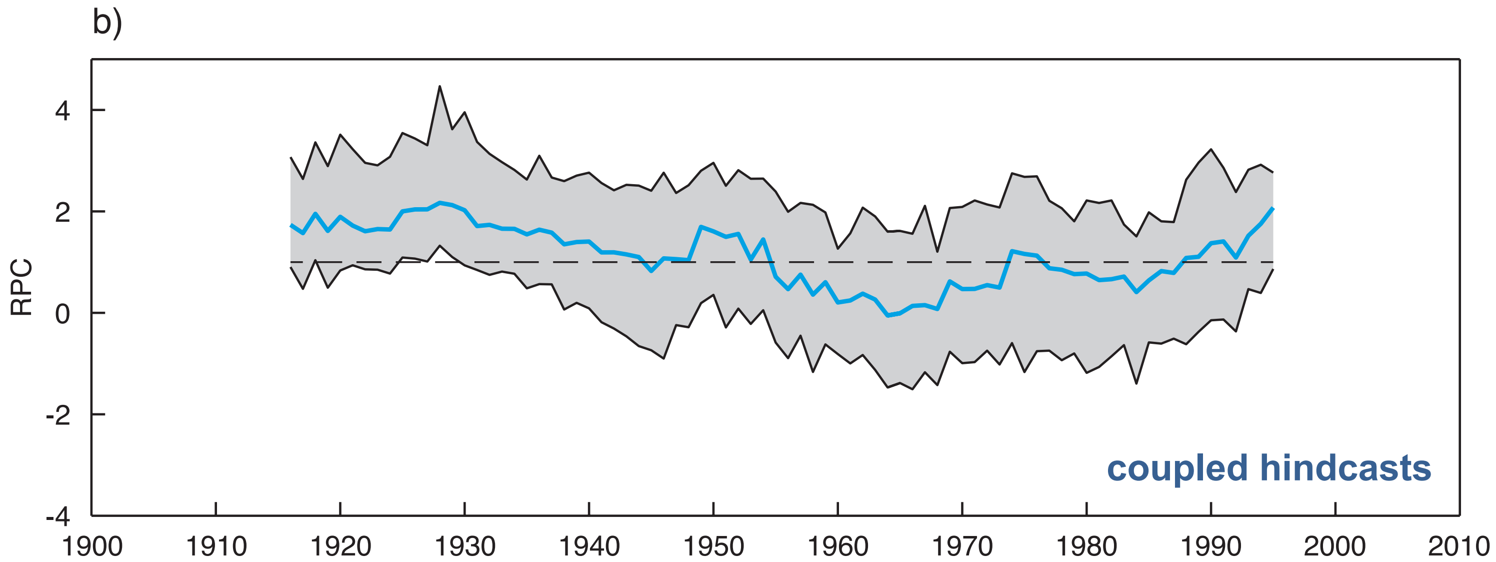
Decadal variability of the Ratio of Predictable Components (RPC) of the NAO in DJF

black dashed lines show the perfect model RPC=1



1901-2010:

RPC=1.05



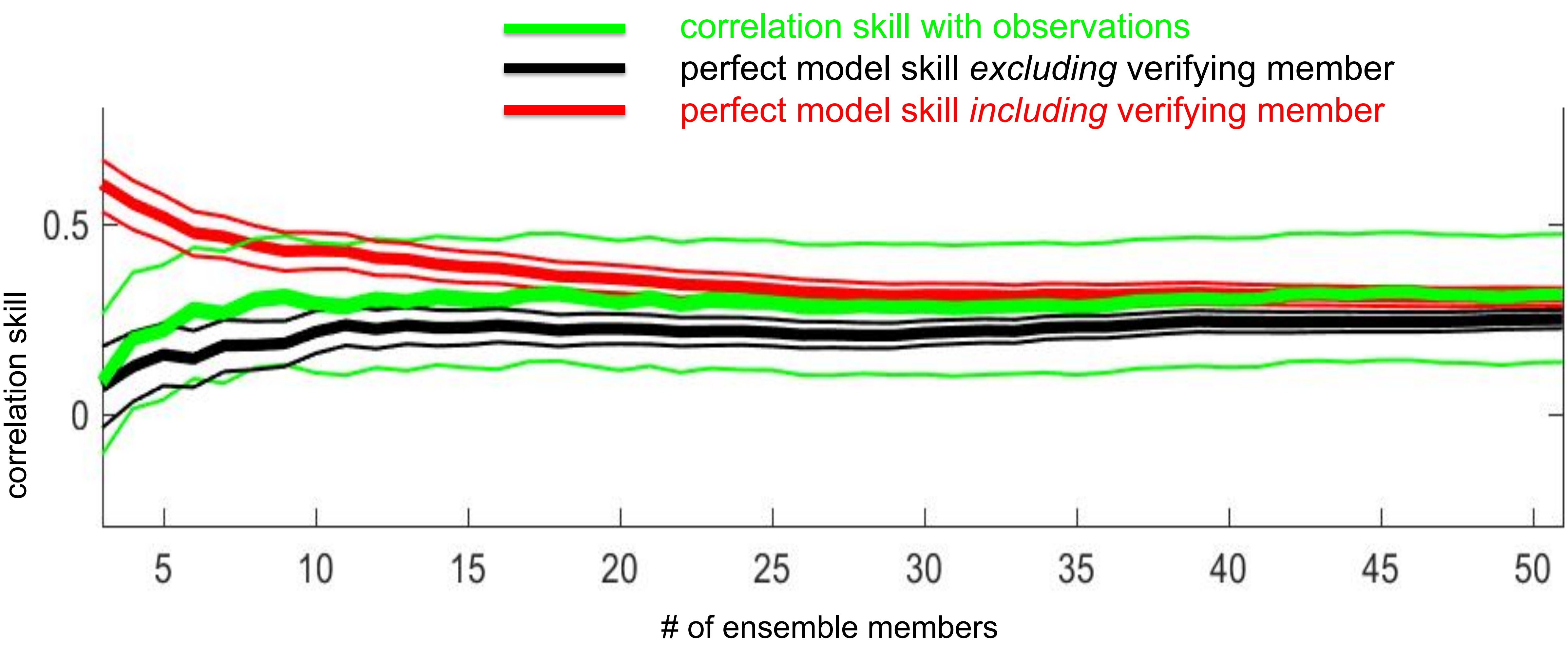
1901-2010:

RPC=1.19

Weisheimer et al. (BAMS 2020)

Signal and noise: what is a perfect model?

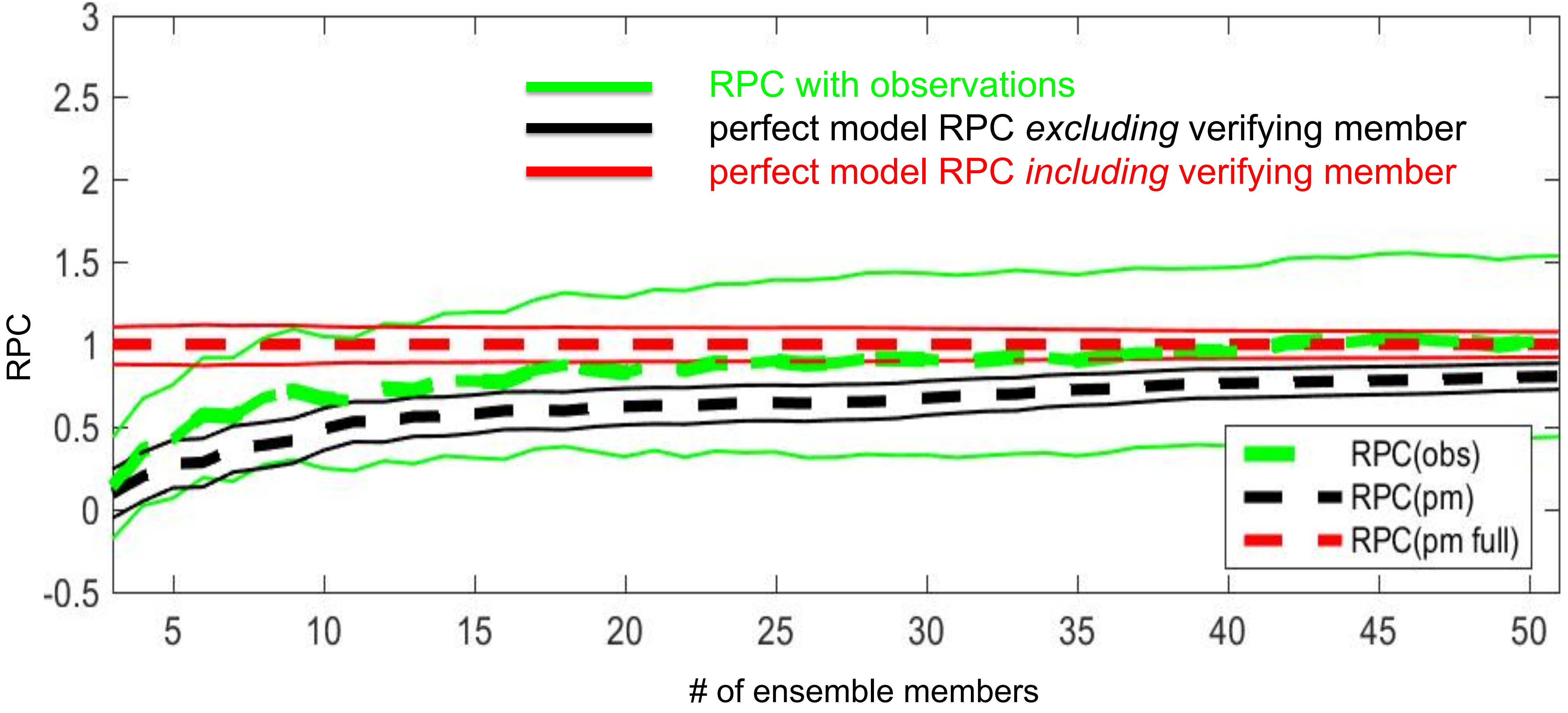
Does a perfect model include or exclude the verifying member?



Weisheimer et al. (QJRMS 2019)

Signal and noise: what is a perfect model?

Does a perfect model include or exclude the verifying member?

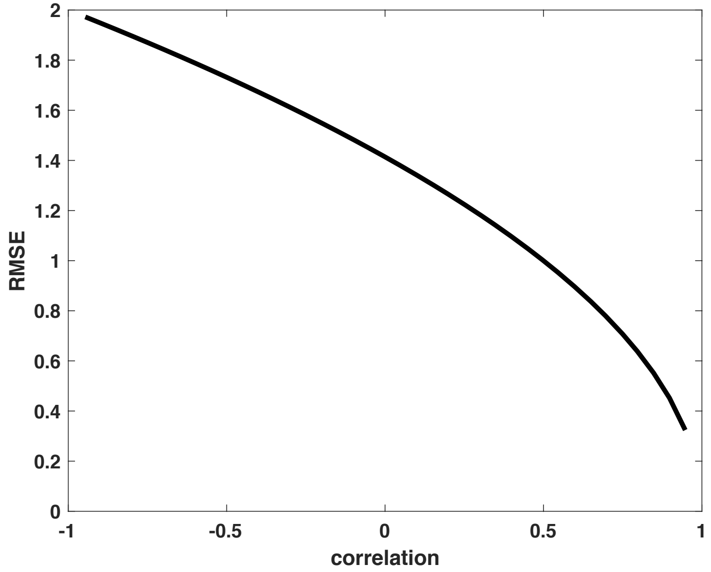


Weisheimer et al. (QJRMS 2019)

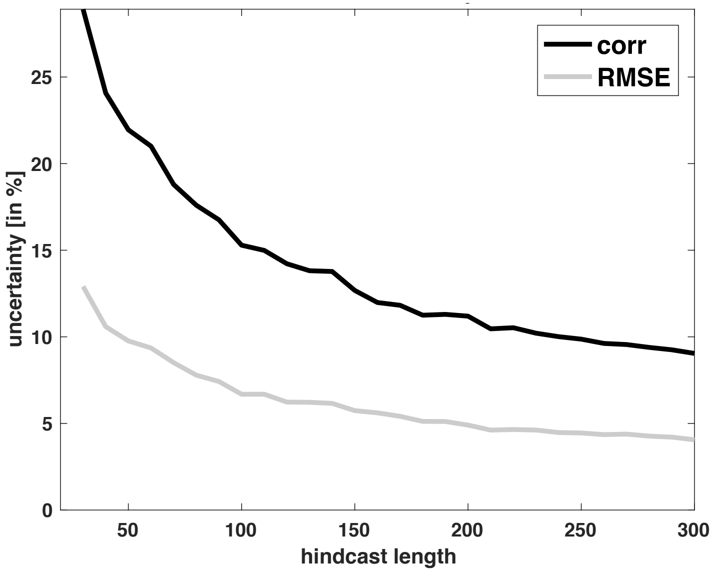
Signal and noise: Robustness of RMSE and correlation skill

- Synthetical long data: “truth” and “ensemble means” with prescribed correlations, n=30,000
- Sub-samples of the “truth” to test robustness for small sample sizes: n=30 … 300
- Uncertainty = normalized standard deviation of distribution of skill measure (correlation or RMSE)

RMSE vs correlation

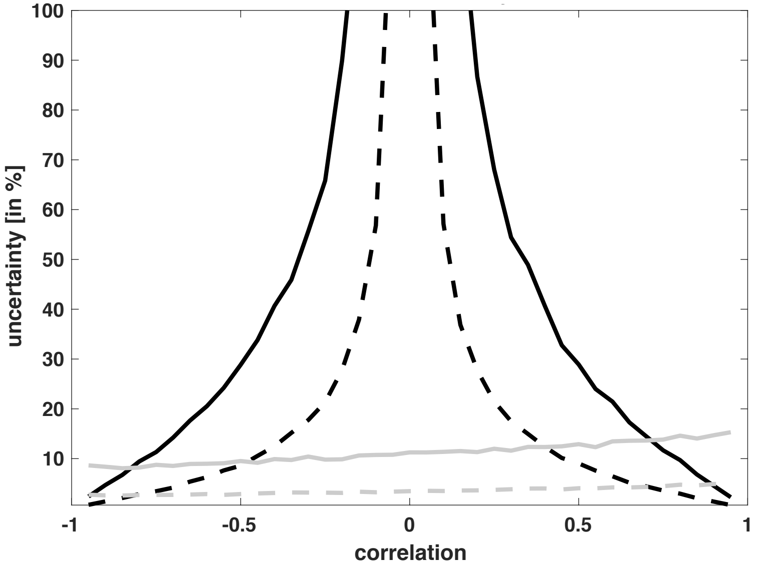


Uncertainty vs sample size



“true” correlation =0.5

Uncertainty vs “true” correlation



correlation RMSE

n=30

n=300

Weisheimer et al., QJRMS (2019)

Conclusions

- New seasonal hindcast data sets from 1901 to 2010 provide a test bed for estimating seasonal predictability during distinct recent climate periods (also useful as test bench for deep learning?)
- Seasonal ENSO forecast skill varies non-monotonically across the Century in coupled and uncoupled hindcasts with similar levels of skill at the beginning and end of 20th Century (role of the observing system?)
- Biennial hindcasts SEAS5-20C (1901-2010) test limits of ENSO predictability out to 24 months
  - Background state changes show complex behaviour (mean, variability)
  - skill drop in spring (barrier) is most sensitive to multi-decadal variability
- Evidence for multi-decadal variability of extratropical winter forecast skill (NAO, PNA) with pronounced drop of skill in mid-Century decades

→ **Short hindcast period are not sufficiently representative for longer-term behaviour (skill, confidence) due to decadal climate variability**

Mid-Century period stands out as an important period on which to test the performance of future seasonal forecast systems

→ **Achieving good forecast skill for recent decades is not sufficient to guarantee similar good performance in the future**

References (I)

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.

Hoskins, B. J. and D. J. Karoly (1981), The steady linear response of a spherical atmosphere to thermal and orographic forcing, *J. Atmos. Sci.*, *38*, 1179–1196.

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.*, doi:10.5194/gmd-2018-228

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, D. MacLeod, D. Befort and T.N. Palmer (2020). Assessing the robustness of multidecadal variability in Northern Hemisphere wintertime seasonal forecast skill *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.3890

Scaife, A.A. and D. Smith (2018). A signal-to-noise paradox in climate science. *npj Clim. Atmos. Sci*. **1**, 28.

References (II)

Shukla, J. (1998). Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science*, **282**, 728-731.

Smith, D.M., Scaife, A.A., Eade, R. *et al.* (2020). North Atlantic climate far more predictable than models imply. *Nature* **583,** 796–800. https://doi.org/10.1038/s41586-020-2525-0

Stockdale, T., S. Johnson, L. Ferranti, M. Balmaseda and S. Briceag (2018). ECMWF's new long-range forecasting system SEAS5. *ECMWF Newsletter*, **154**, doi:10.21957/tsb6n1

Wallace, J.M. and D.S. Gutzler (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109**, 784-812.

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

Weisheimer, A., D. Decremer, D. MacLeod, C. O'Reilly, T. Stockdale, S. Johnson and T.N. Palmer (2019). How confident are predictability estimates of the winter North Atlantic Oscillation? *Q. J. R. Meteorol. Soc.*,**145**:S1, 140-159, doi:10.1002/qj.3446

Weisheimer, D. Befort, D. MacLeod, T.N. Palmer, C. O'Reilly and K. Strommen (2020). Seasonal forecasts of the 20th Century *Bull. Amer. Meteor. Soc.*, **101** (8): E1413–E1426, doi:10.1175/BAMS-D-19-0019.1