Predictability: It's all about the Signal and the Noise . . .

Prashant.D.Sardeshmukh@noaa.gov
CIRES, University of Colorado and Physical Sciences Laboratory/NOAA
Thanks to: **Aaron Wang, Gil Compo, Cecile Penland, Matt Newman, Sang-ik Shin, Joe Barsugli**

What is the "true" or "potential" predictability of the Earth System ?

1. **How do we <u>estimate</u> it ?**

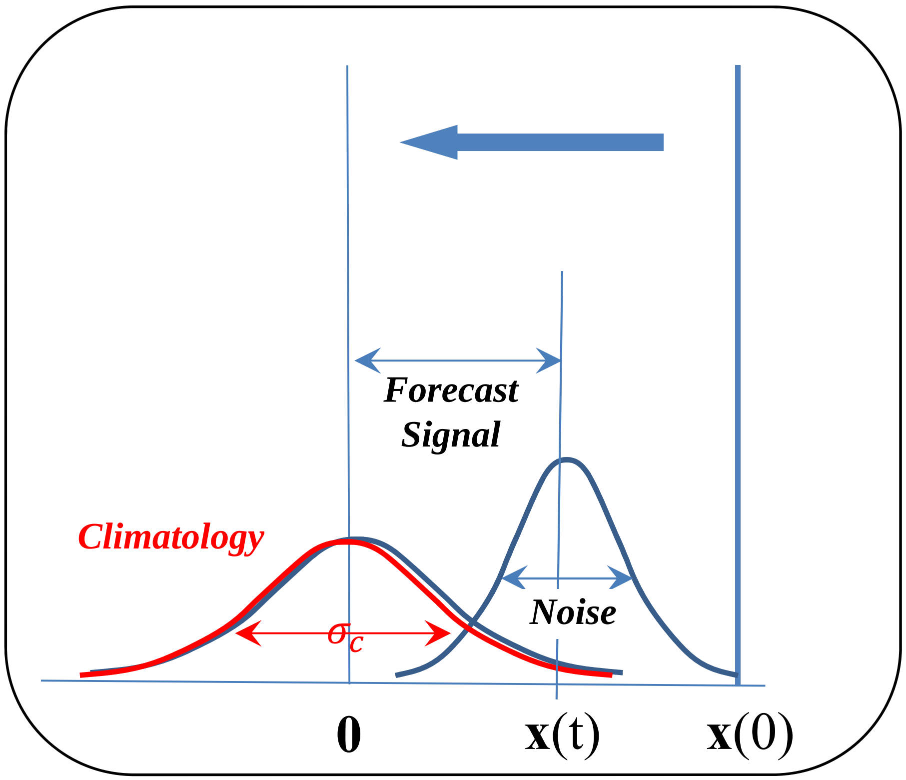
2. **How do we <u>attain</u> it ?**

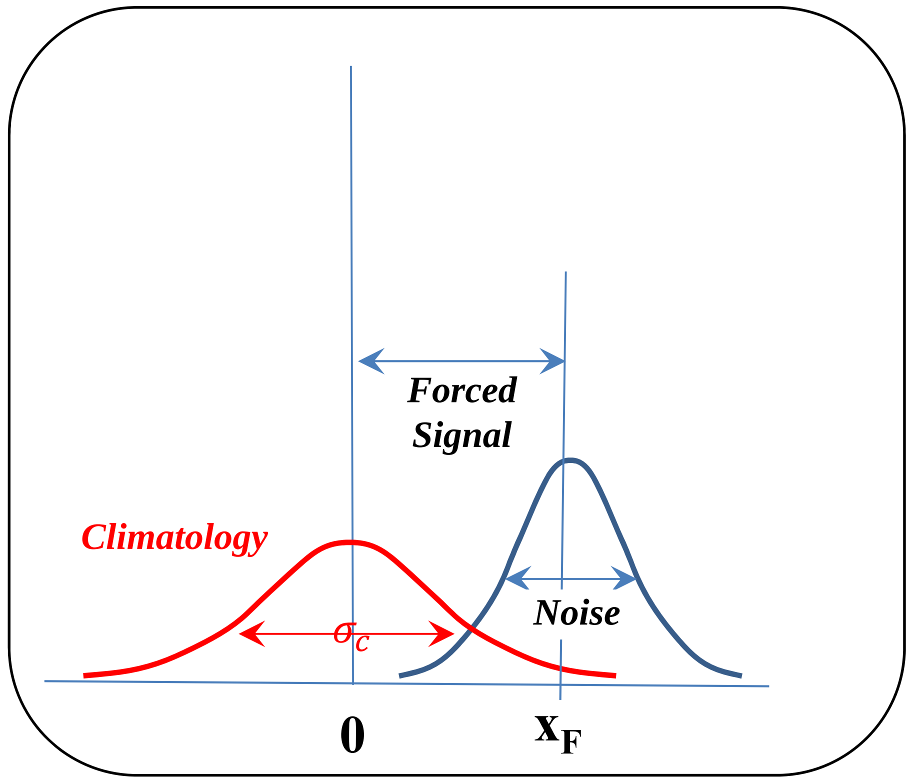
3. **How do we <u>exploit</u> it ?**

Predictability of the "First" and "Second" kind in a chaotic system

<u>First kind</u> : associated with evolution from known **initial conditions**

<u>Second kind</u>: associated with response to *predictable* slowly varying **forcing**

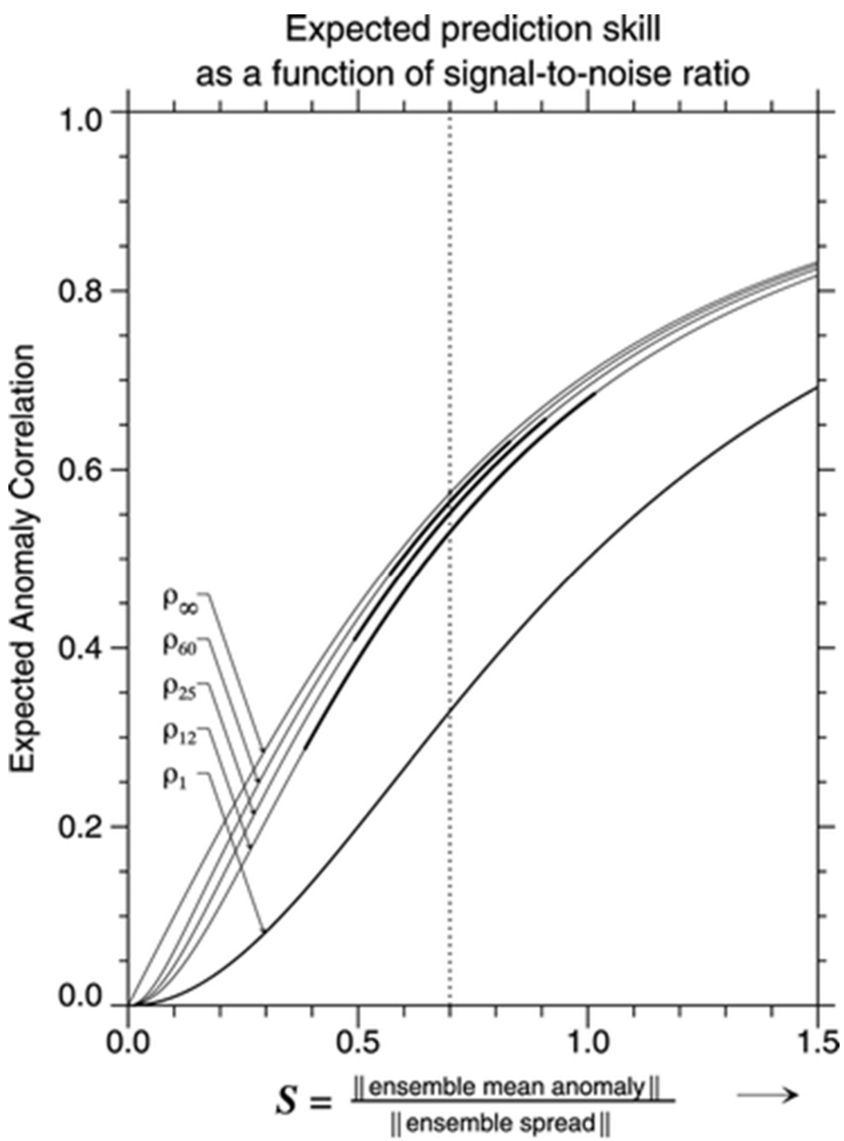




In both cases, predictability is determined by the *Signal-to-Noise ratio S*

There is predictability if *S* is not zero (more broadly, if the forecast *pdf* differs from the climatological *pdf*)

We define predictability here as the expected correlation $\boldsymbol{\rho_\infty}$ *of observed and infinite-member ensemble-mean anomaly forecasts made using a perfect model with perfect initial conditions.*



The maximum expected correlation and the associated minimum r.m.s. error are given by:

$$\rho_\infty^2 = \frac{S^2}{1+S^2} \quad \text{and} \quad \epsilon_\infty^2 = (1-\rho_\infty^2)\,\sigma_c^2$$

The correlation is smaller if finite *n*-member ensembles are used, and its estimation is also more uncertain (**thickened** portions of curves)

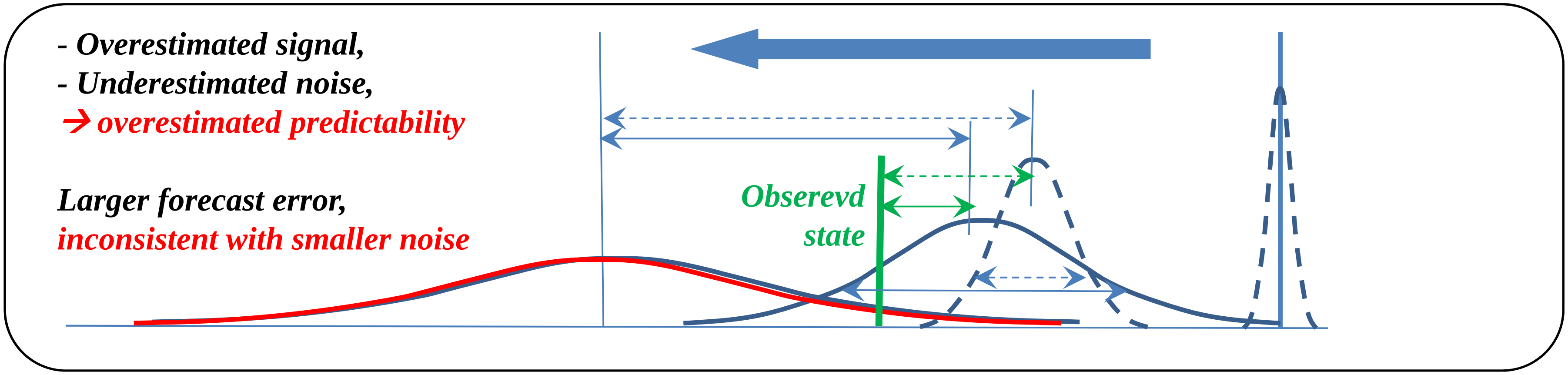
Predictability in any forecasting context may be assessed by specifying the relevant estimated value of *S* on this plot.

***S* may be estimated using ensemble forecasts, but its value may be compromised by errors in estimating the signal as well as the noise.**

Compo and Sardesmukh 2004

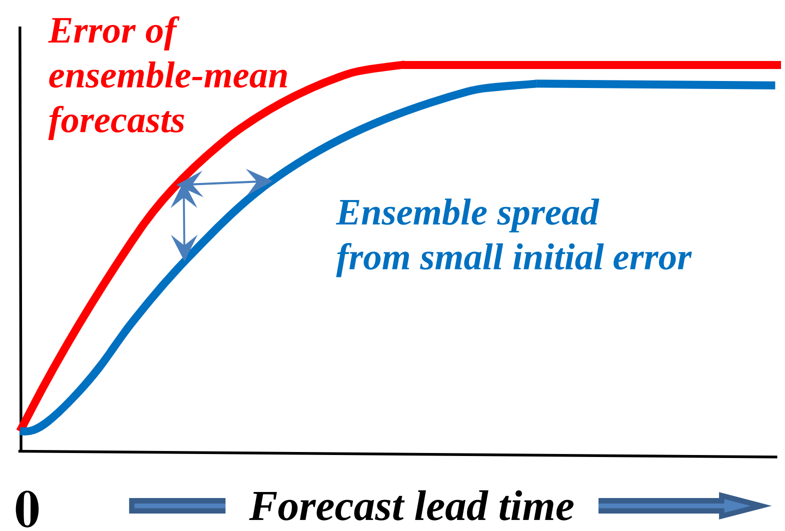
The forecast *pdf* differs from the "true" *pdf* because of model errors and initial errors

This has consequences, not just for forecast errors but also for estimating predictability



If the forecast *pdf* differs from the "true" *pdf*, the forecast error and ensemble spread (i.e. the noise) growth curves will not match.

Lorenz suggested using the gap between these curves to quantify the potential for forecast improvement.



*An increasingly popular approach to **reducing the gap** between the forecast error and spread curves is to introduce additional stochastic terms in a model's equations*

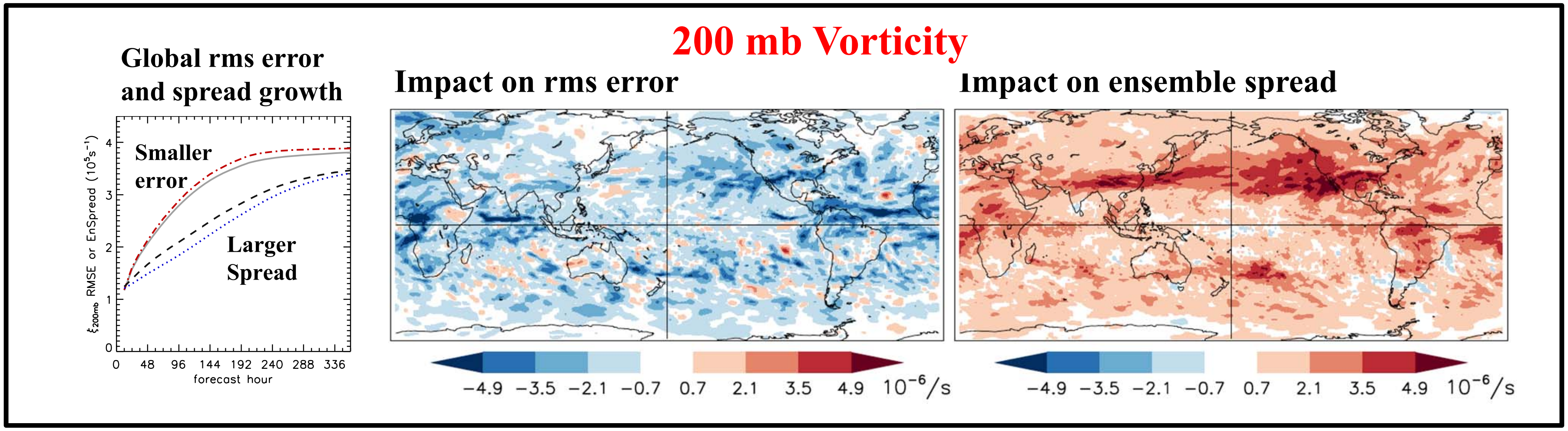
$$\frac{dx}{dt} = \underset{\textit{resolved}}{A(x)} + \underset{\textit{parameterized}}{P(x)} + \underset{\textit{unparameterized}}{R}$$

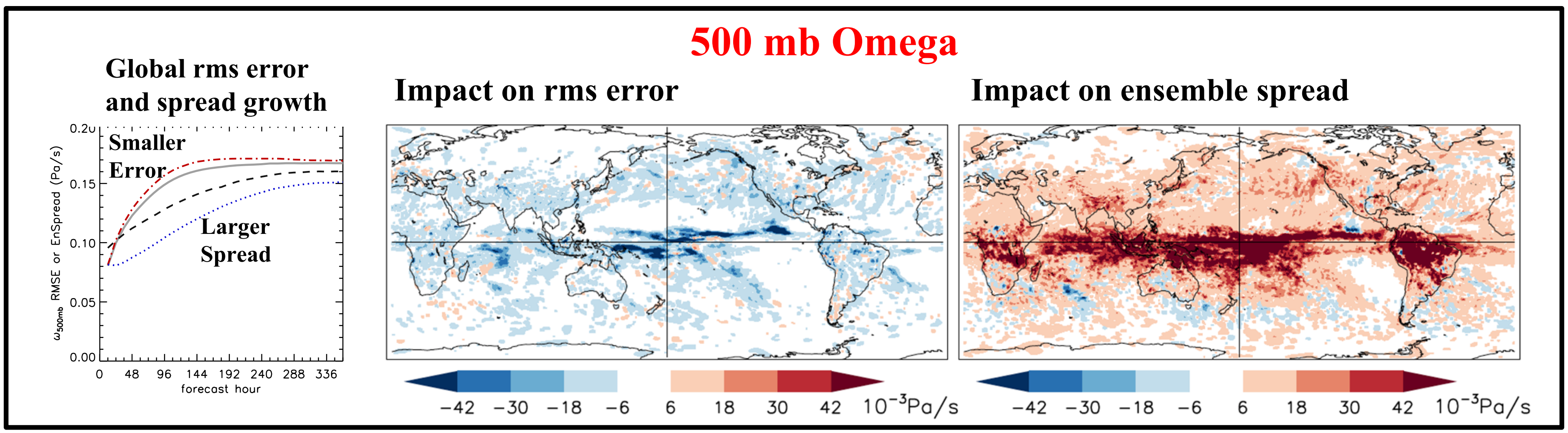
$$\approx A(x) + (1+r)P(x)$$

Specifying R ~ r P(x) at every model grid point, where r is a random number between +1 and -1 and is spatially correlated over ~ 500 km is an effective way to account for chaotic atmospheric physics in models

*Even this crude approach has proven successful at **reducing the gap** in several weather and climate prediction contexts, not only by increasing the ensemble spread but also by decreasing the forecast error.*

For example, for ***subseasonal*** (Day 15) predictions, stochastic parameterizations of the form $(\mathbf{1}+\boldsymbol{r})\boldsymbol{P}(\boldsymbol{x})$ in a T254 version of the NCEP/GFS model leads to both a **reduction of the rms error** of the ensemble-mean forecasts and an **increase of the ensemble spread,**





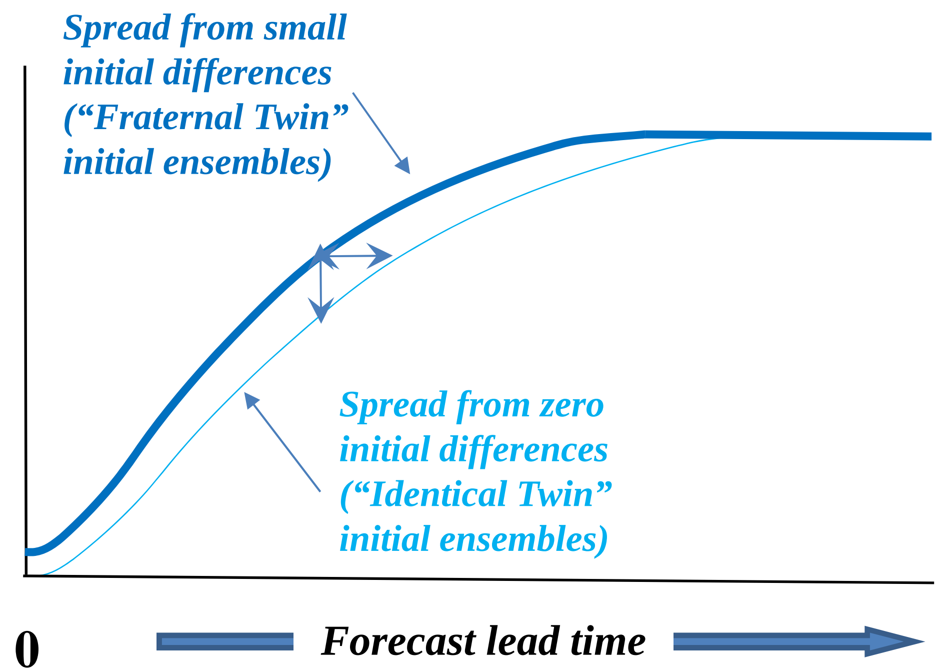
These results are for 80-member ensemble forecasts for 80 separate forecast cases in Jan-Mar 2016 (*Sardeshmukh, Wang, Compo, and Penland, 2020*)

*The stochastically perturbed NCEP model allows us to perform a "**Dream" Calculation**:*

Assessing the Lorenz Predictability of Global Weather

*by examining the spread of 6400 forecast pairs starting from identical initial conditions (that one may call **Identical Twin ensembles** a la Lorenz's classic predictability experiment) with the spread of the 80-member ensemble forecasts starting from initial analysis ensembles (that one may call **Fraternal Twin ensembles**) such as shown on the previous slide*

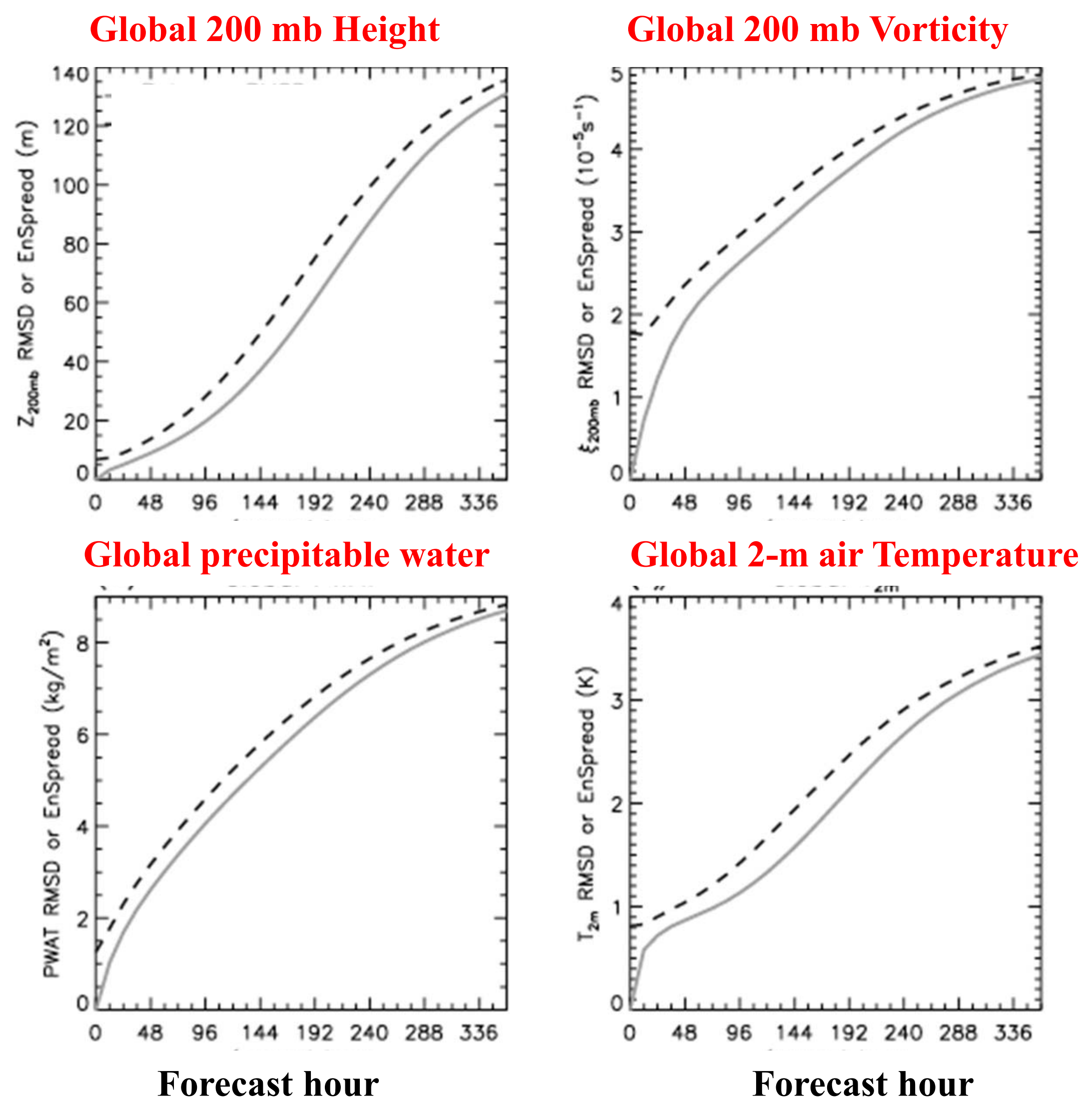
The gap between the curves provides an estimate of the potential gain in weather predictability from eliminating initial errors



The size of the "gap" obtained suggests that one can improve weather prediction skill by ~ 1 day by eliminating initial errors on even large resolved (~ 50 km) scales

Dashed curves: Spread of "Fraternal Twin" ensembles

Solid Curves : Spread of "Identical Twin" ensembles



Sardeshmukh, Wang, Compo 2020

On <u>subseasonal and longer</u> time scales, it is useful to approximate the predicable anomaly dynamics as ***linear in a low-dimensional space*** *and the unpredictable chaotic nonlinear dynamics as a* ***stochastic noise forcing***

$$\frac{dx}{dt} \approx Lx + b\eta_1 + (Ex + g)\eta_2$$

This approximation adequately captures subseasonal anomaly dynamics, including the non-Gaussianity of subseasonal anomalies

$$\approx Lx + B\eta$$

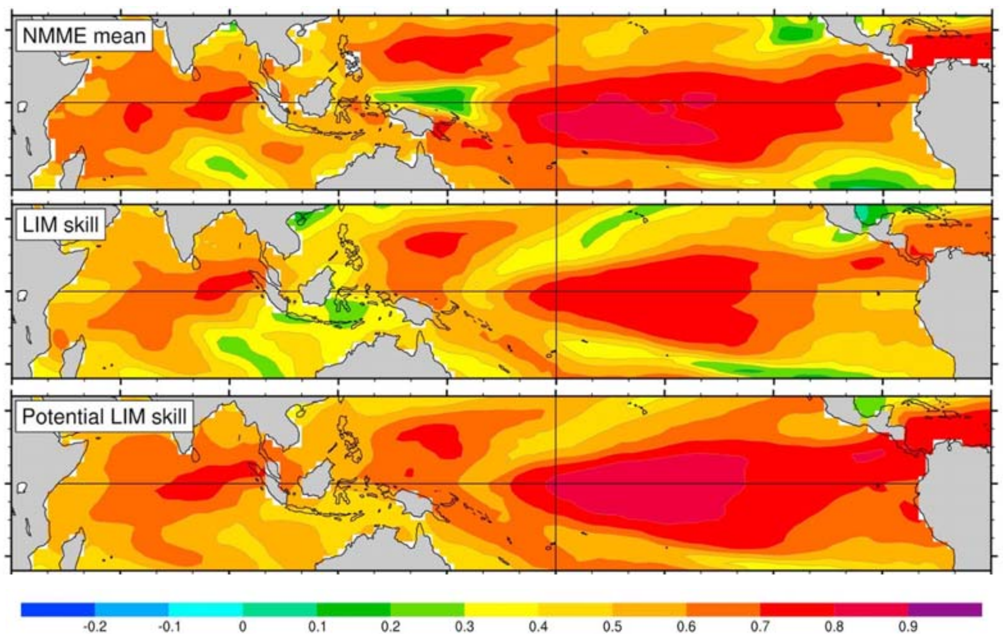
This approximation adequately captures seasonal and longer time scale anomaly dynamics

The dynamical feedback matrix ***L*** *and the stochastic forcing amplitude matrix* ***B*** *can be estimated directly from data through Linear Inverse Modeling (One may legitimately call this "Machine Learning" !)*

This has proven to be a remarkably good approximation. But even more important for our purposes here, it allows one to investigate predictability by <u>explicitly</u> identifying the L terms with the forecast ***<u>signal</u>*** *and the B terms with the forecast* ***<u>noise</u>****.*

For example, for ***seasonal*** tropical SST predictions, a Low-Order (28-component) model of the form $\boldsymbol{\frac{dx}{dt} = Lx + B\eta}$, where $\boldsymbol{L}$ and $\boldsymbol{B}$ are estimated through Linear Inverse Modeling (*Penland and Sardeshmukh 1995*) has very similar skill to that of the models used in the operational National Multi-Model Ensemble (NMME) system.

SST anomaly correlation skill at Month 6



NMME-mean skill

LIM skill

Potential LIM skill ρ_∞

The state vector x in the LIM comprises SST and SSH anomalies

From Newman and Sardeshmukh 2017

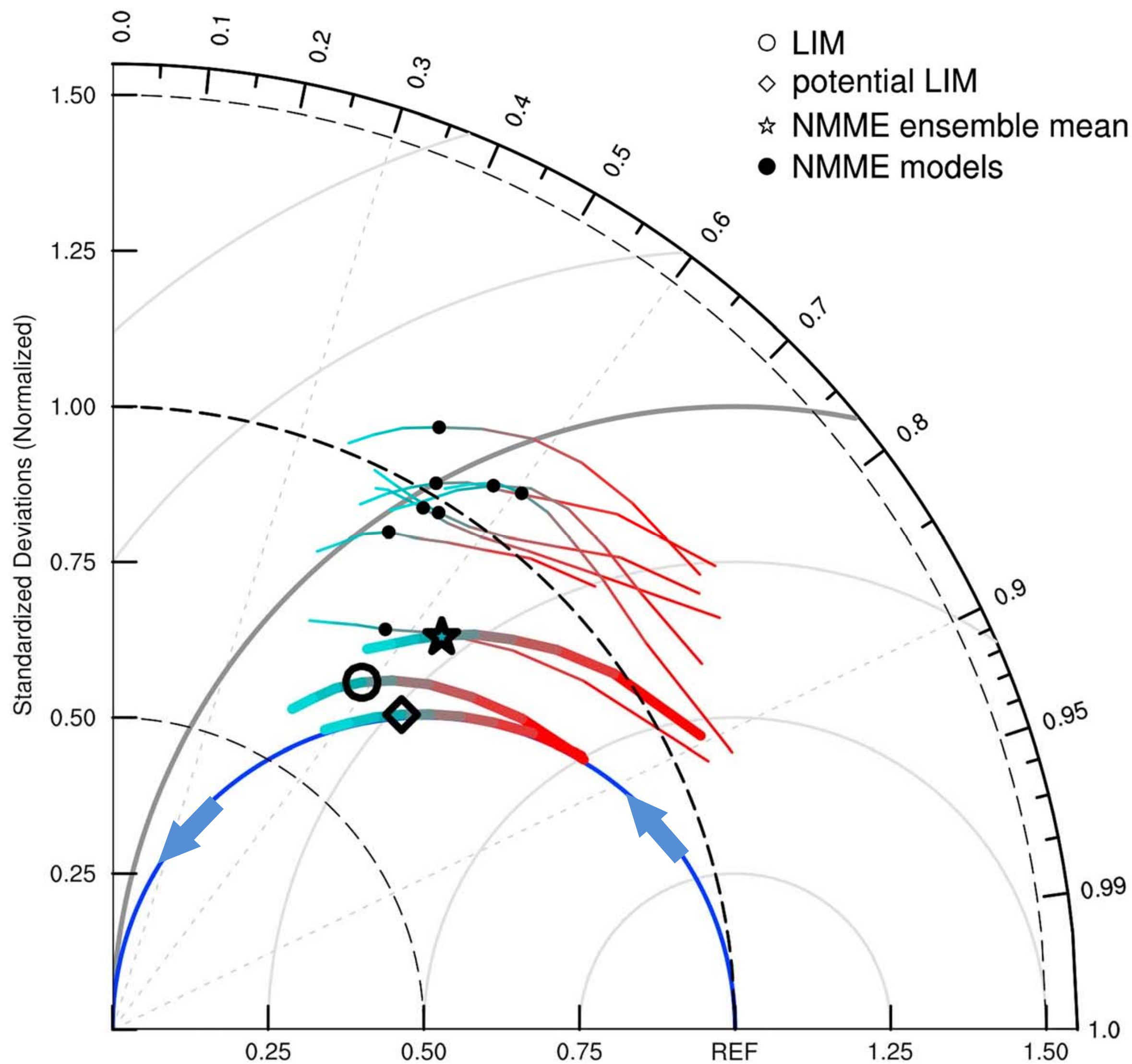
The LIM skill is higher than that of all the individual NMME models, and is comparable to the NMME-mean skill as well as the "potential" skill (blue semicircle)

Taylor Diagram showing decay of SST forecast skill of the 8 NMME models, the NMME mean, and the LIM, from Month 1 (RED end of curve) to Month 9 (**cyan** end of curve), with the black symbol showing the Month 6 skill.

The **blue semicircle** shows the "perfect model" forecast skill trajectory consistent with the relationship

$$\epsilon_\infty^2 = 1 - \rho_\infty^2$$

we discussed earlier.

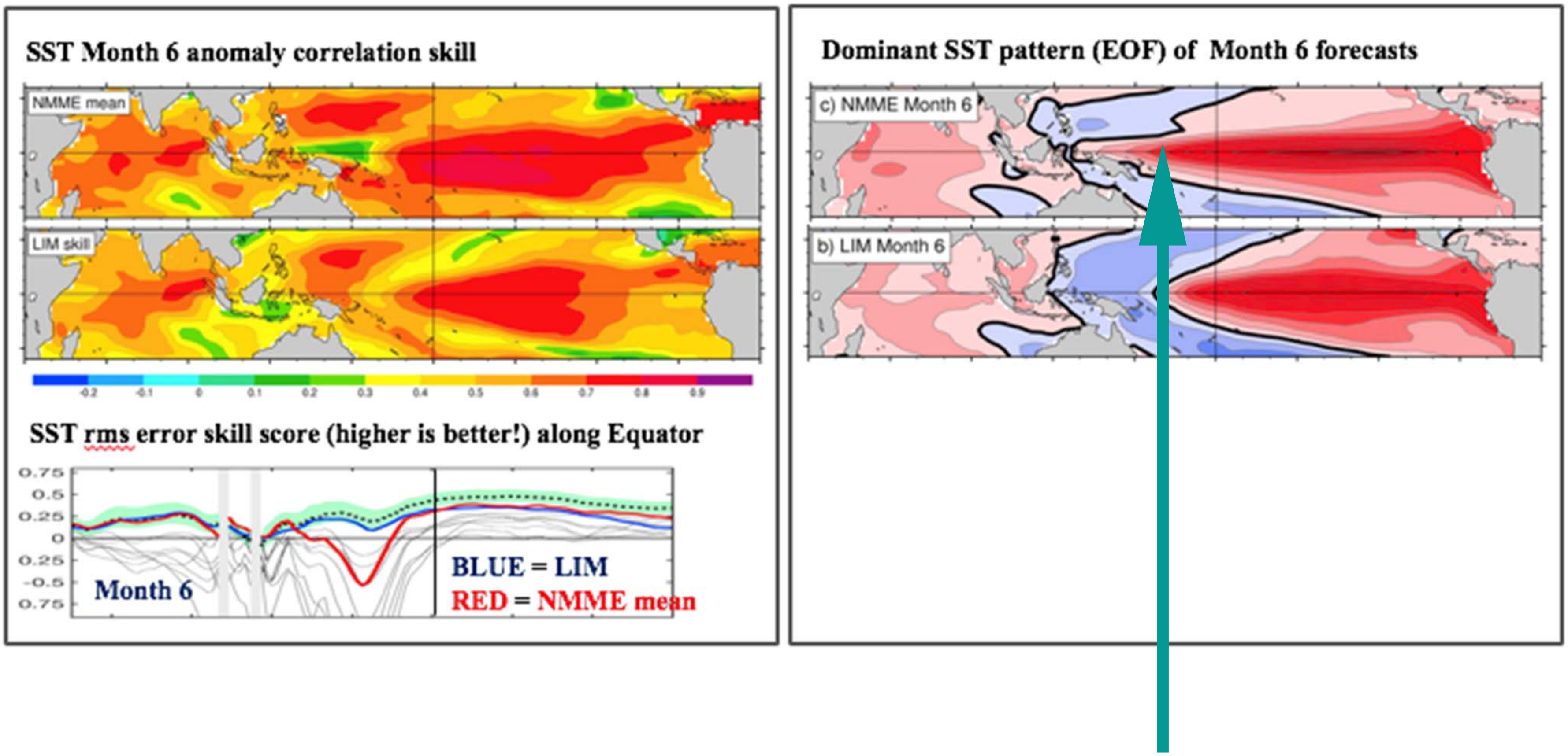


Any point on this plot has the anomaly correlation as its angular coordinate, the normalized forecast amplitude as its radial coordinate, and the normalized rms error as the distance between the point and "REF"

The main area where the LIM clearly outperforms the NMME models is the western Pacific.

This is basically because ENSO extends too far west in the NMME models.

The LIM doesn't have this problem. (*Newman and Sardeshmukh* 2017)

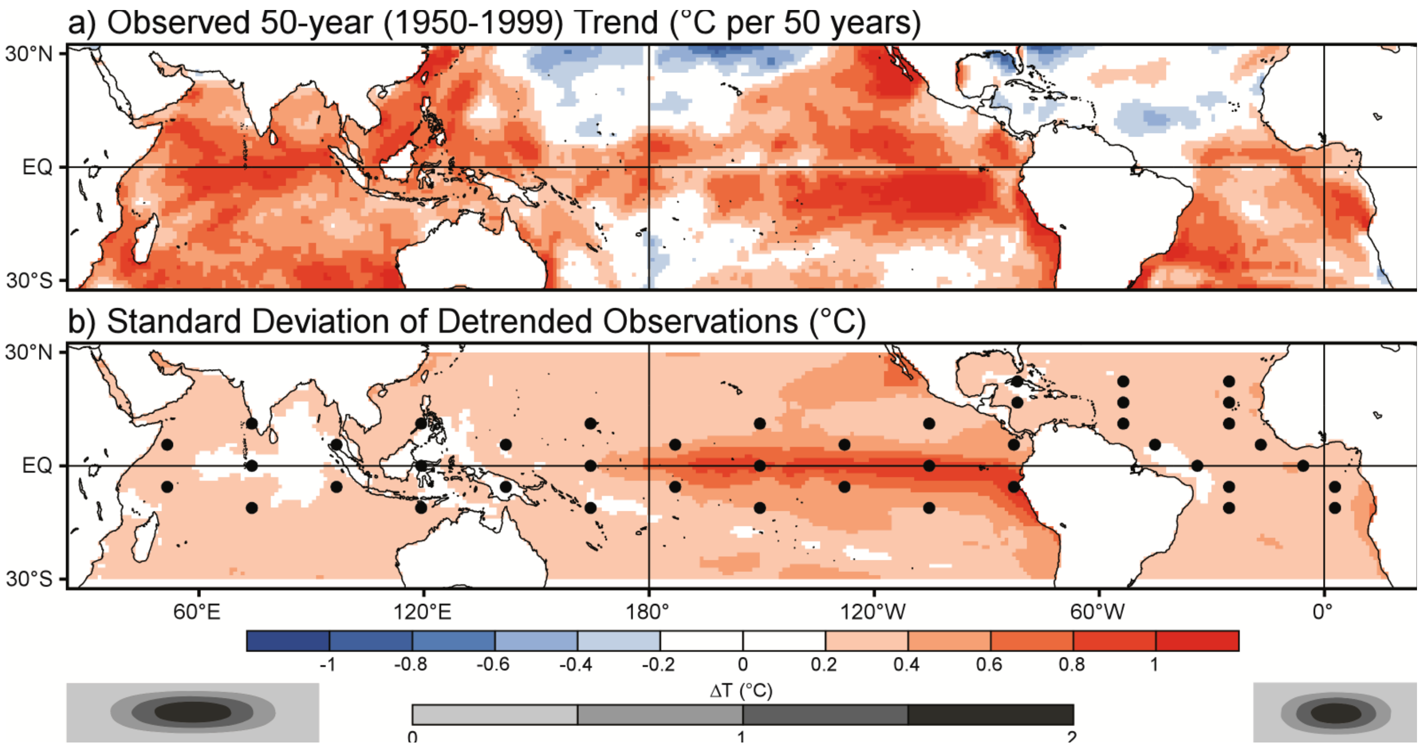


Unfortunately, the global climate is particularly sensitive to SST anomalies in this area.

Fig 10

To estimate global sensitivities to tropical SST anomalies in different tropical areas, we determined an atmospheric GCM's global responses to 43 localized tropical SST anomaly "patches"

(Results are shown below for NCAR/CCM3; results for ECHAM5 are very similar)



Color shading shows the observed HadISST

50-yr (1950-99) SST trends in this panel

and

the standard deviation of detrended seasonal SST anomalies in this panel.

Specifically, we determined the model's seasonally varying ensemble-mean responses to each one of the 43 localized $\pm$ 2/3°C SST anomaly "patches" prescribed at the indicated locations.

Barsugli, Shin, Sardeshmukh

Just two response patterns account for most of the 43 annual responses to the 43 patches

These are the combined response EOFs of the 500 mb height, land temperature, and land precipitation responses

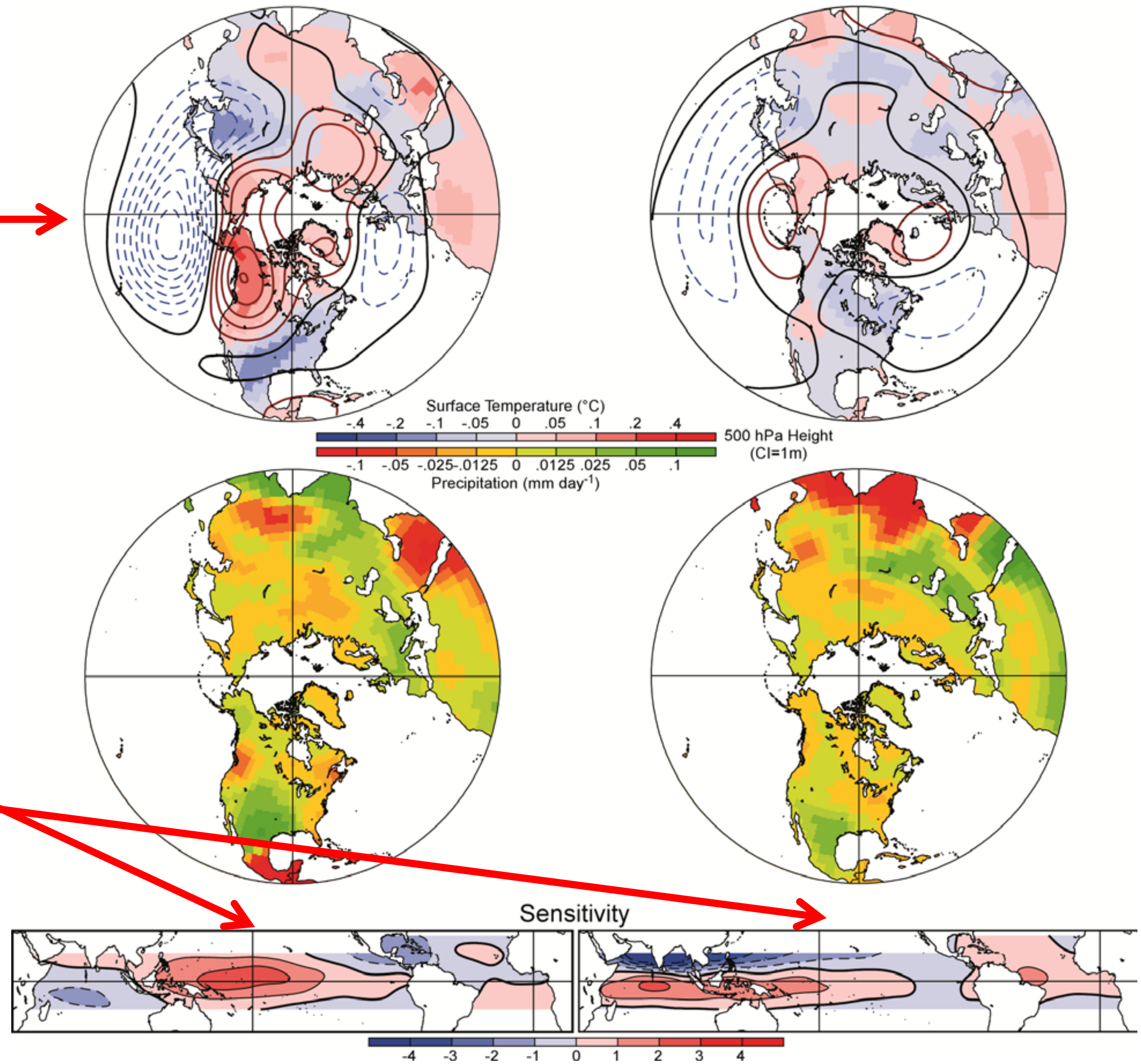
(Optimal response patterns)

And these are the amplitudes with which the response EOFs are excited by the SST patches at different locations

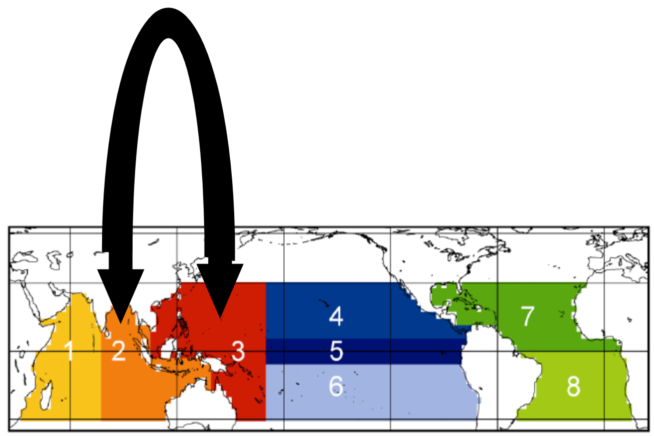
(Optimal SST forcing patterns, also interpretable as Sensitivity patterns)

Note the large sensitivity to the western Pacific SSTs

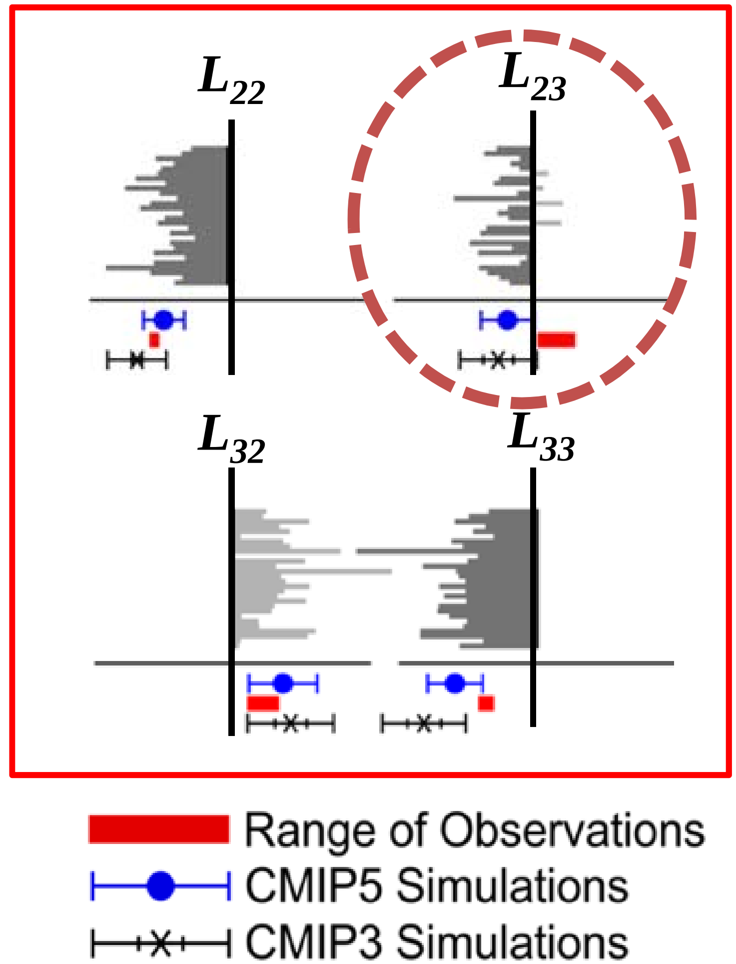
EOF-1 (36 %) **EOF-2 (18 %)**



Local and remote SST feedbacks on SSTs in the Indo-Pacific Warm Pool
estimated from SST covariances among 8 tropical regions using a Linear Inverse Model (LIM)



L_{ij} = **Effect of SST in Region *j* on SST in Region *i***



Almost all the CMIP models have a **<u>wrong-signed feedback</u>** of the Western Pacific ocean SSTs on the Eastern Indian Ocean SSTs (L_{23}).

This is a major concern, with global implications for estimating and attaining Earth System predictability from subseasonal to climate change scales.

SUMMARY

Predictability: It's all about the Signal and the Noise . . .

1. The signal-to-noise ratio S in any prediction context determines an upper bound on potential predictability. The actual anomaly correlation skill cannot exceed $\boldsymbol{\rho}_{\infty}$, which is a simple function of S.

2. One can approach $\boldsymbol{\rho}_{\infty}$ by using large forecast ensembles (*much larger* ensembles if S is small, which it usually is many long-range prediction contexts)

3. Since predictability is estimated using imperfect models, model estimates of S are compromised by errors in both model signals and model noise.

4. The noise in most ensemble forecasting systems is underestimated. This leads to an overestimation of predictability. The problem can be partly remedied by implementing stochastic parameterizations of chaotic physics.

5. Model errors lead to incorrect estimation of forecast signals. The errors in the tropical Indo-Pacific warm pool region are a major concern in this regard, for both estimating and attaining Earth System predictability globally from subseasonal to climate change scales.