Metrics and diagnostics for seasonal forecasts evaluation

MedCOF 2016
Rome

Panos Athanasiadis, CMCC

email: panos.athanasiadis@cmcc.it

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

What is a Seasonal Prediction System?

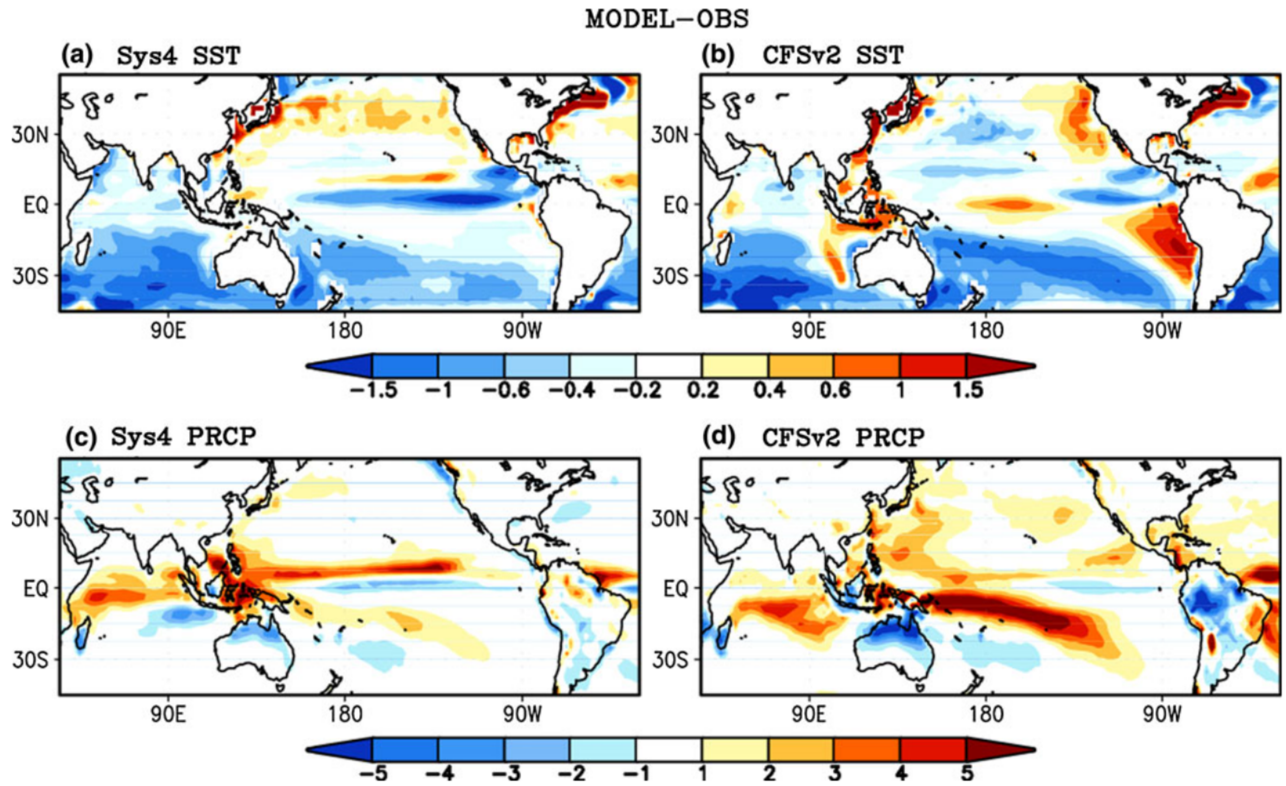
- *Basically, it consists of a climate model (CGCM) run at high resolution and initialized realistically in all its components (atmosphere, ocean, sea ice, snow cover, soil moisture, etc). A large ensemble of forecasts is fundamental.*

- *Predictive skill has quite a different meaning for weather forecasts (~days), seasonal forecasts (~months), decadal forecasts (~years) and climate forecasts (~decades or more).*

- *In seasonal forecasting, skill represents the capability of the system to predict monthly mean, or seasonal mean anomalies of meteorological fields (e.g. Tmax, Z500) or indices (NAO, ENSO, frequency of blocking and extreme events) and the likelihood of these exceeding certain thresholds.*

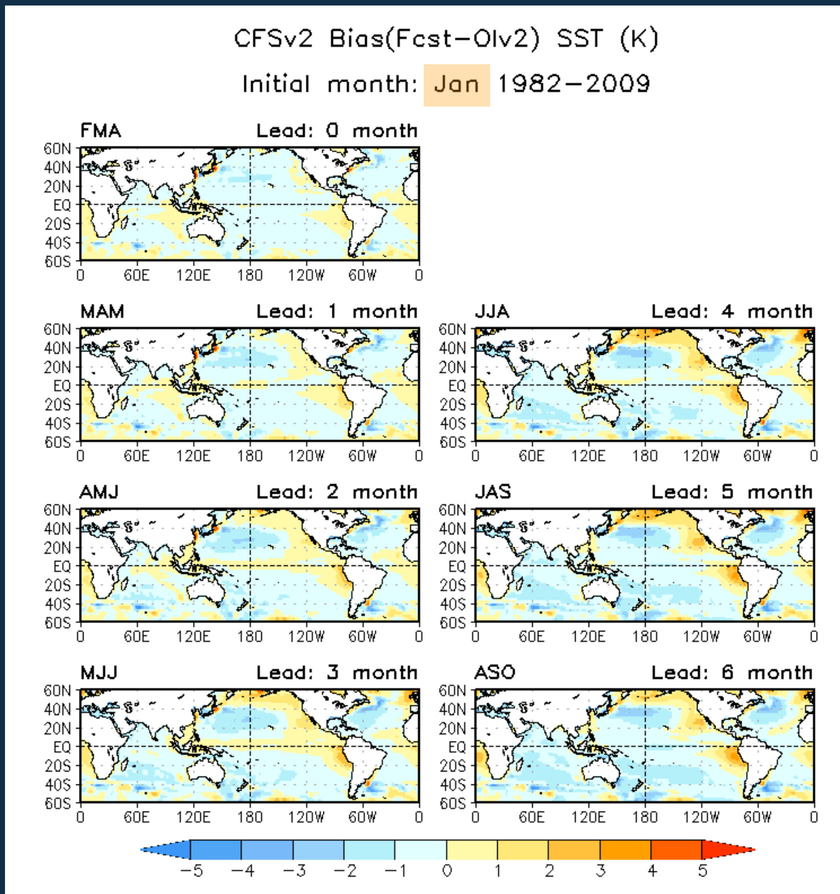
- *A realistic representation of the observed mean climate, seasonal cycle, variance at different timescales, extremes, teleconnections and key physical processes is a prerequisite for good skill.*

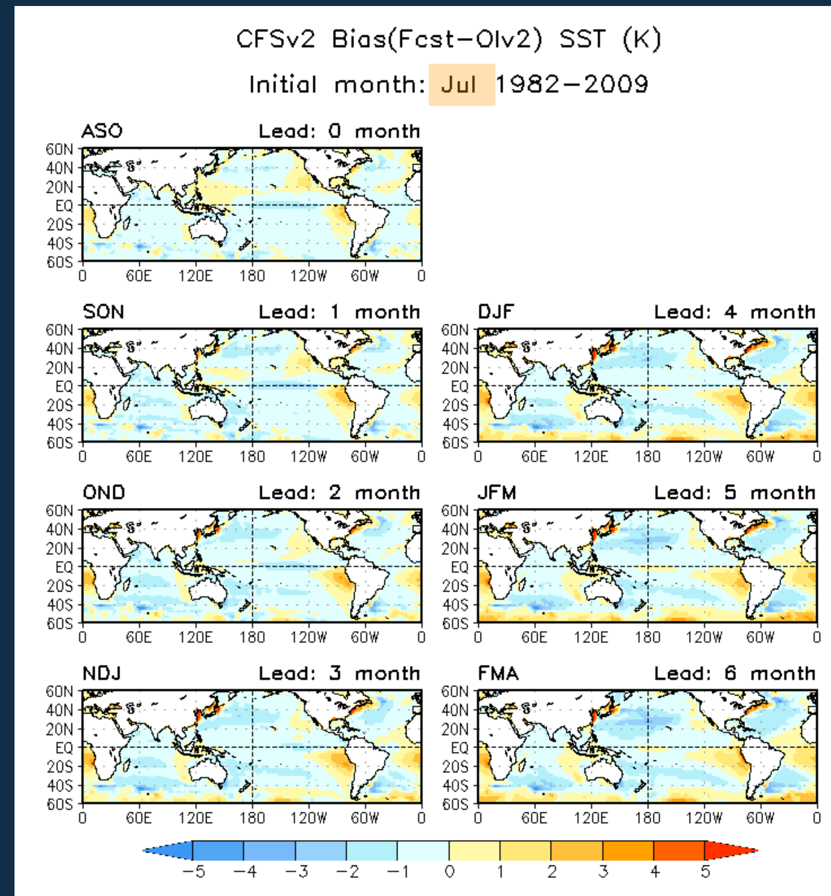
Models have their own equilibrium climate



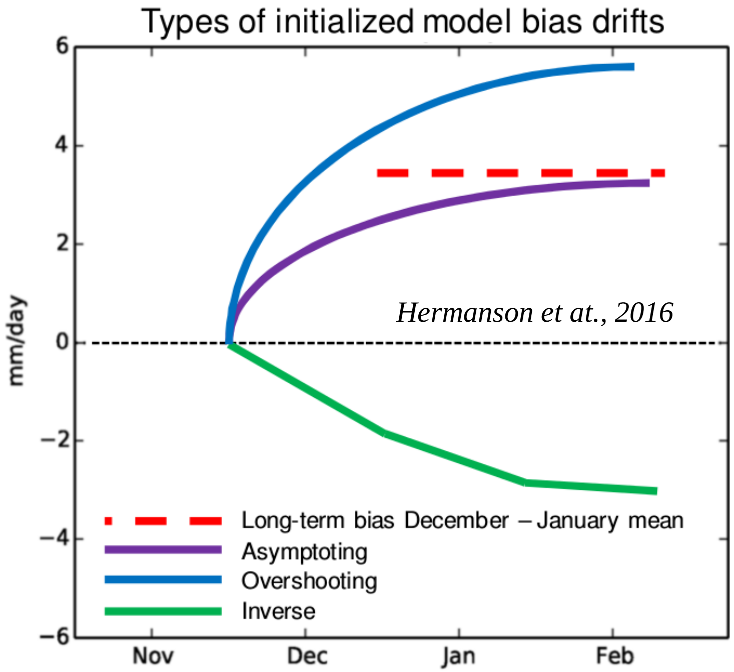
Kim et at., 2012

Biases depend on lead time…



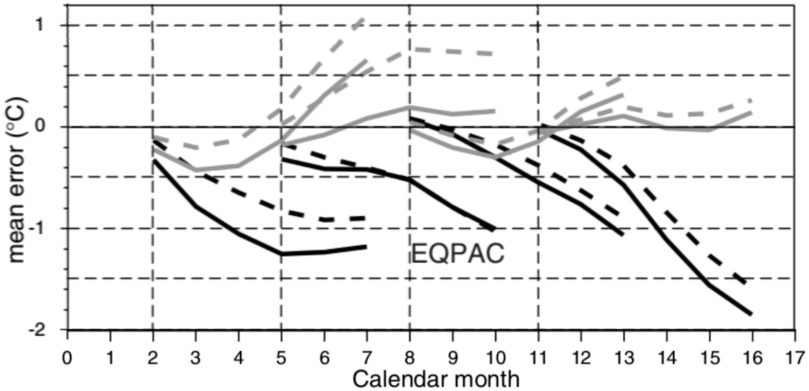


This is called model drift

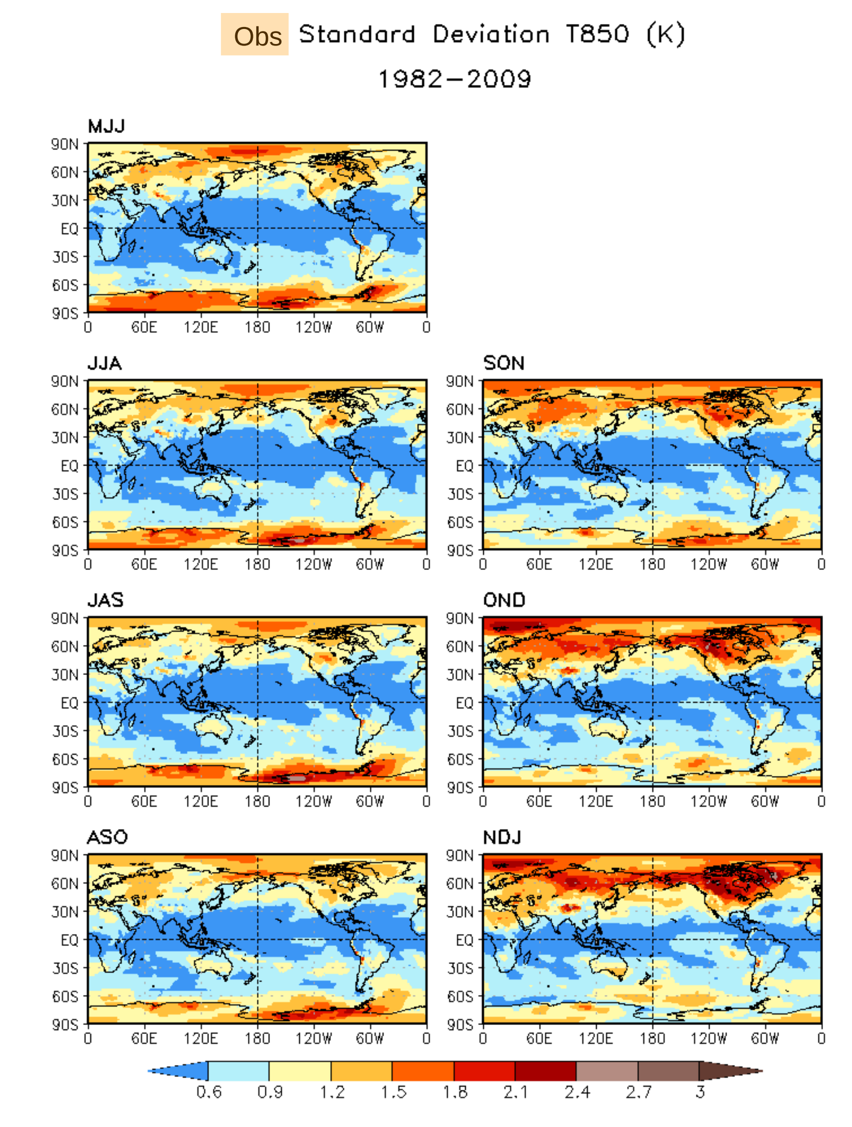


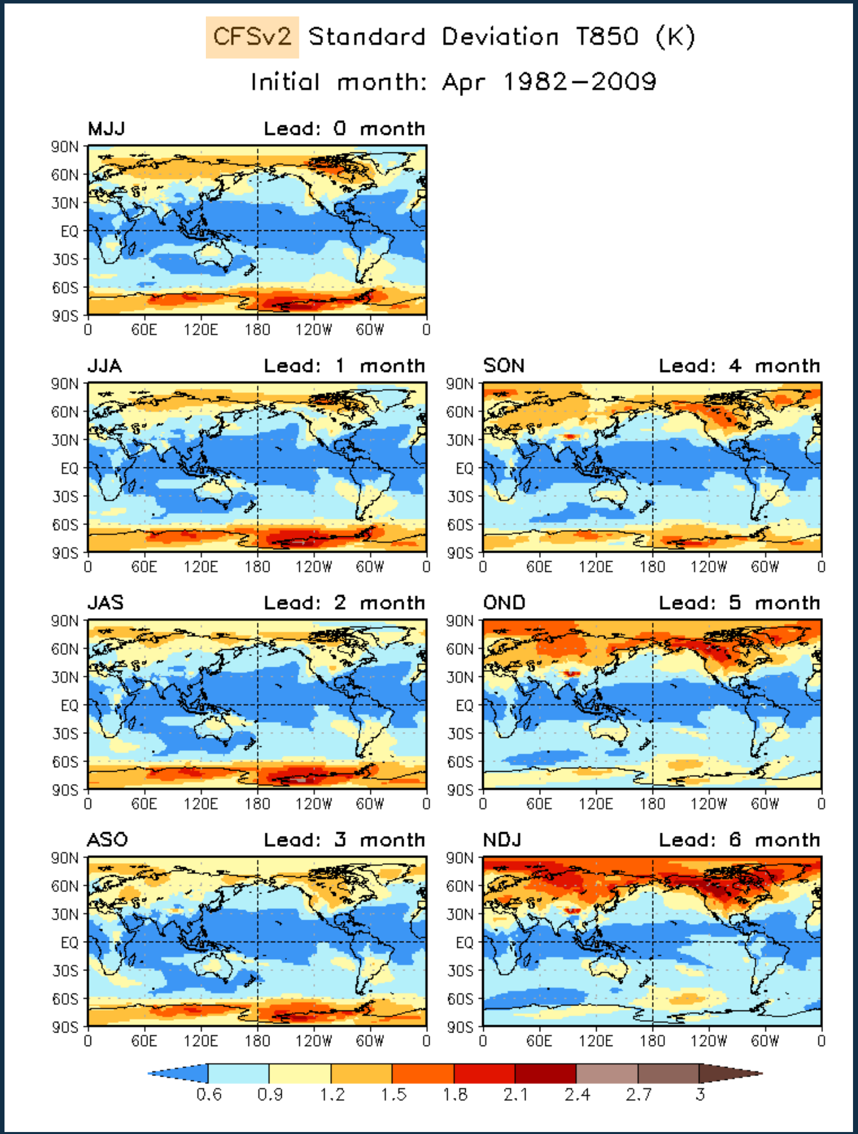
Schematic of the types of drifts encountered in the two seasonal forecast systems. The red dashed line represents the bias in a spun-up control integration using the same model. The hindcasts are initialized 1 November (and 1 May) for at least 15 hindcast years with at least 8 ensemble members. The drifts represent the average development of the bias over all hindcasts and ensemble members. The type of drift is diagnosed from the December – February mean bias. Asymptoting drift is of the same sign and smaller than the long-term bias. Overshooting drift is the same sign and larger than the long-term bias. Inverse drift is of the opposite sign to the long-term bias. Figure 2 shows the drifts we found.

Lazar et at., 2005

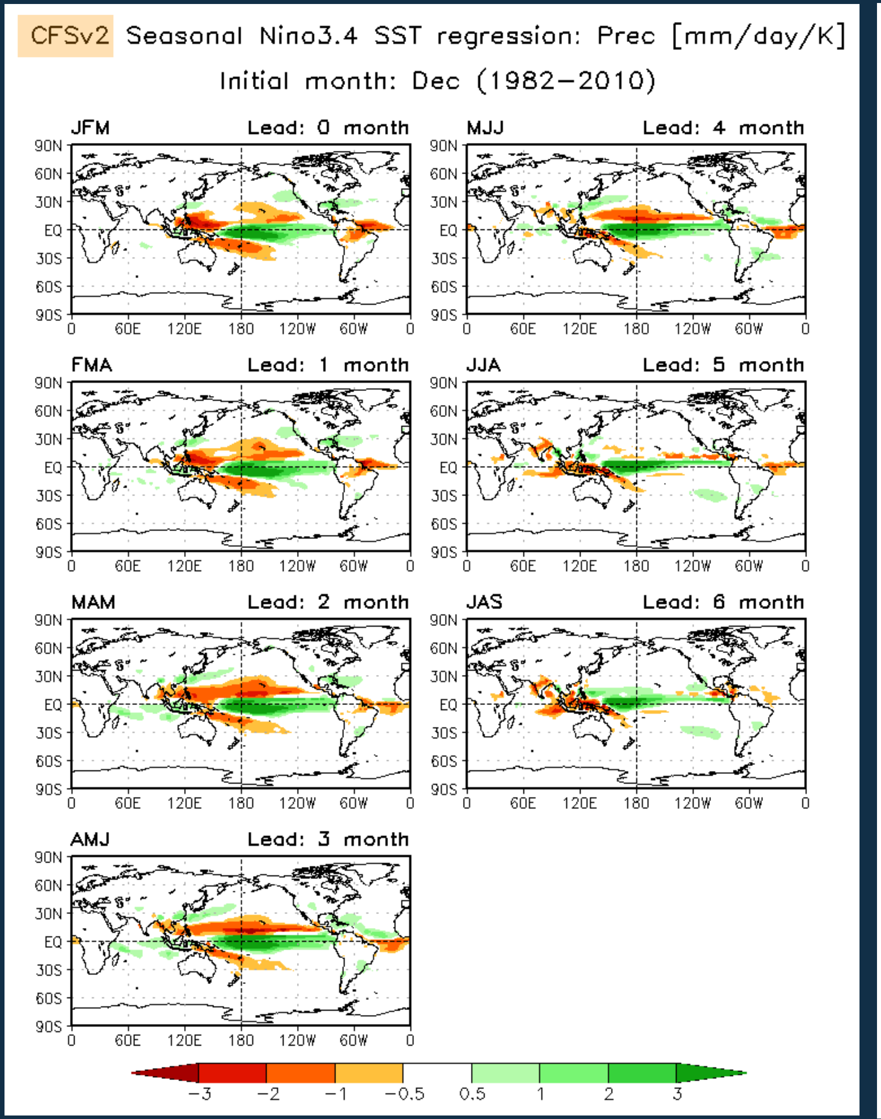


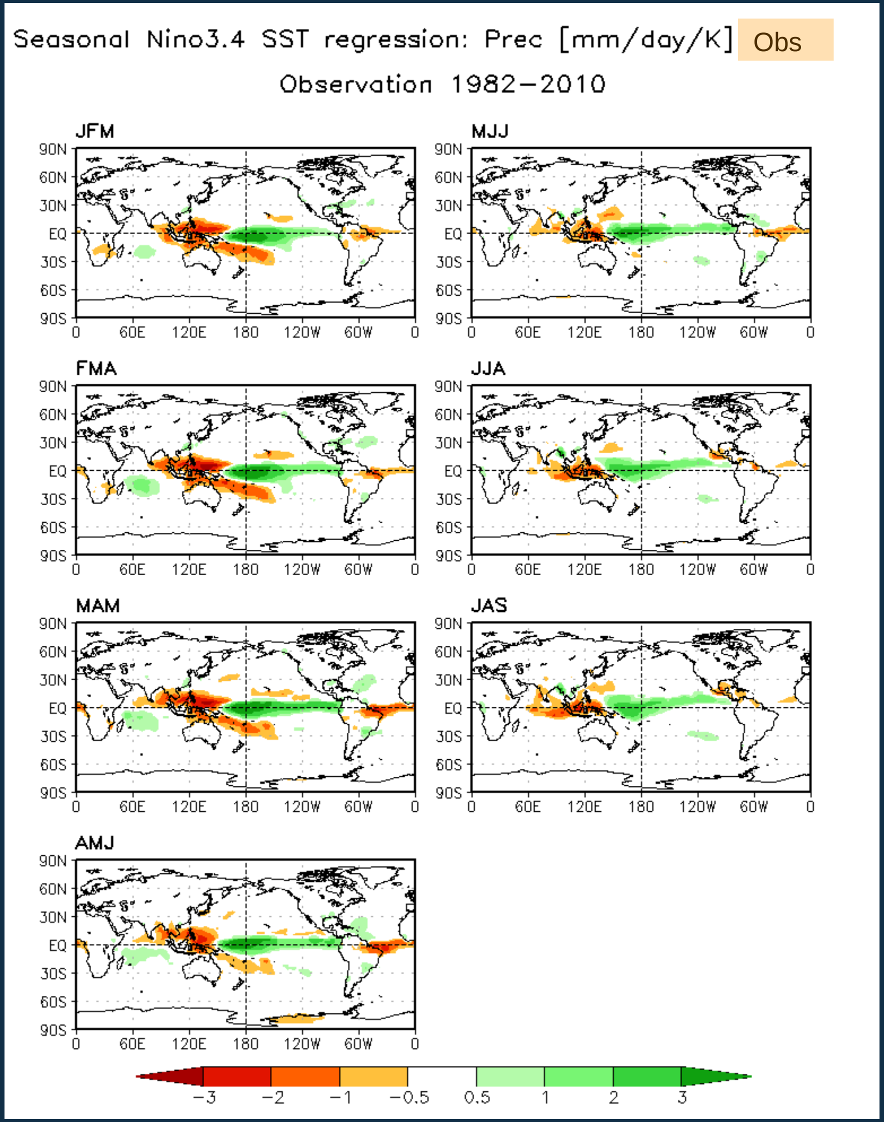
Representation of variability



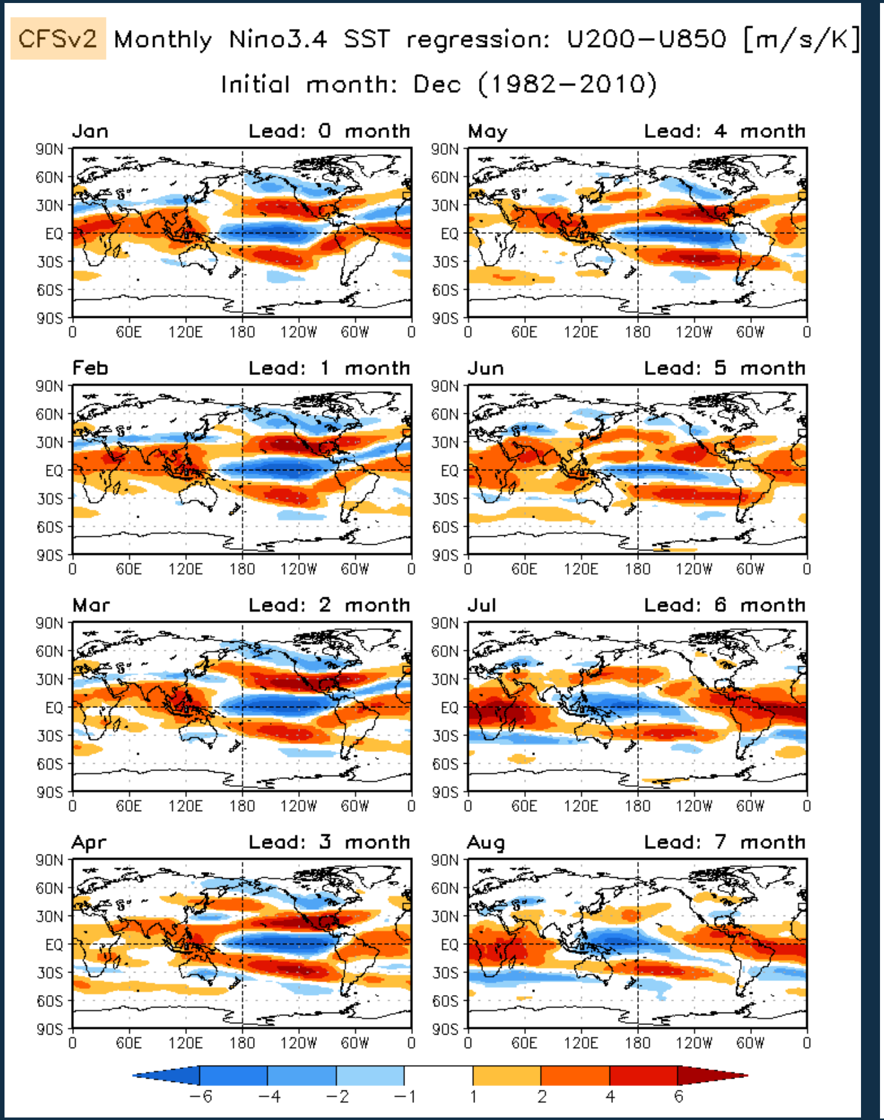


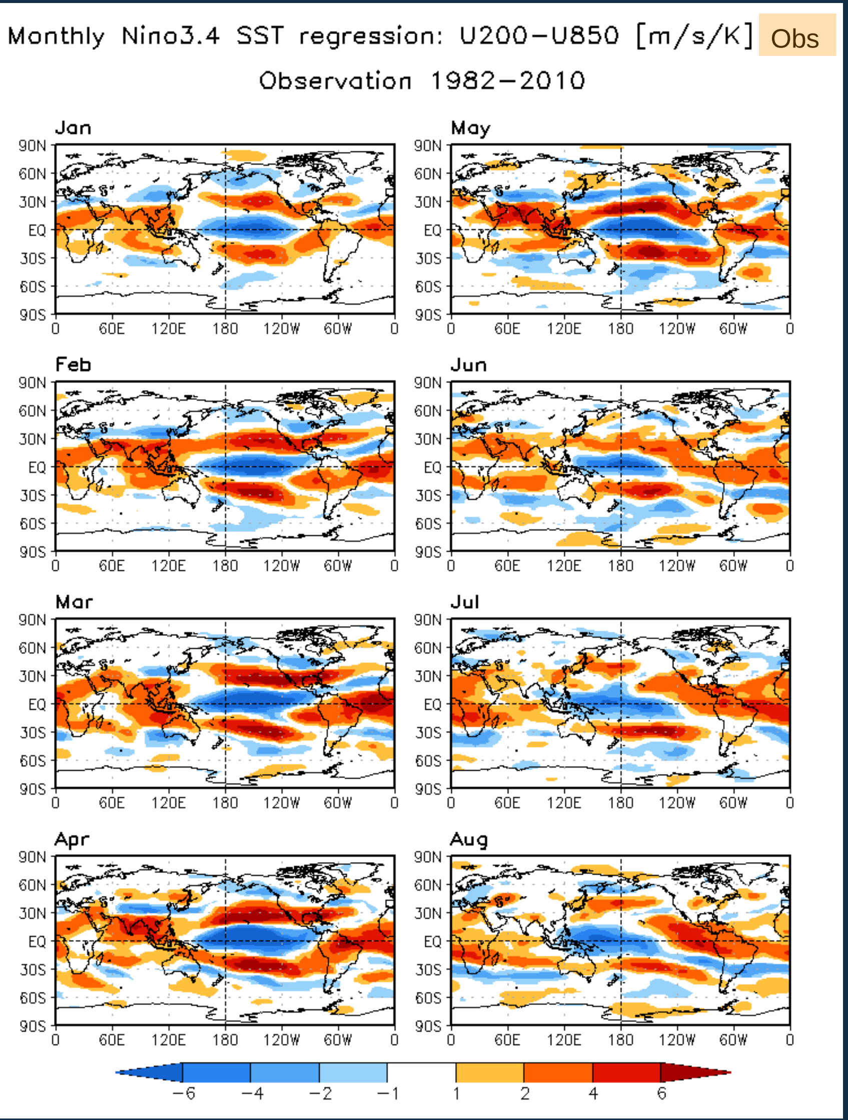
Representation of Teleconnections



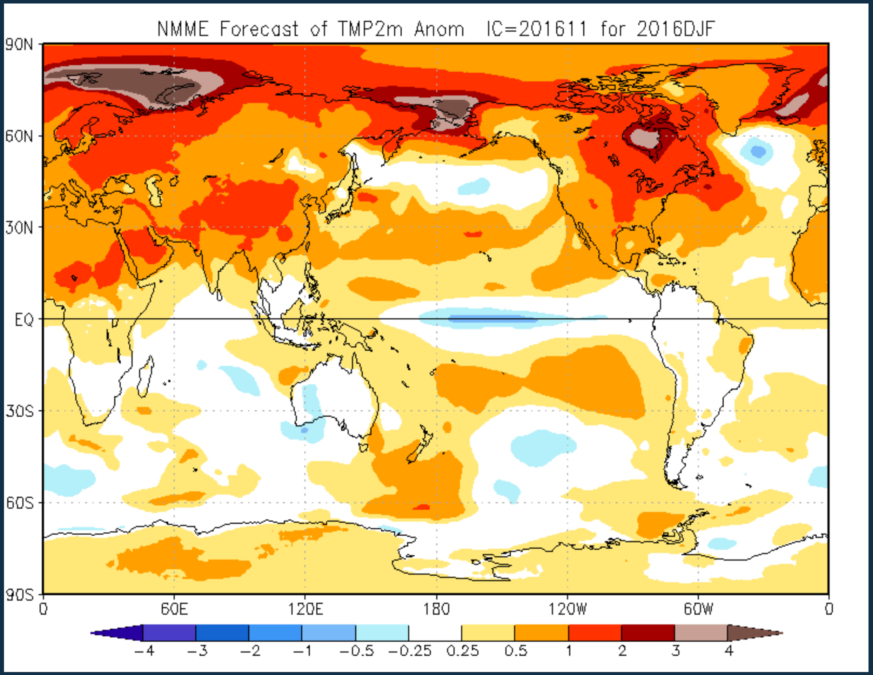


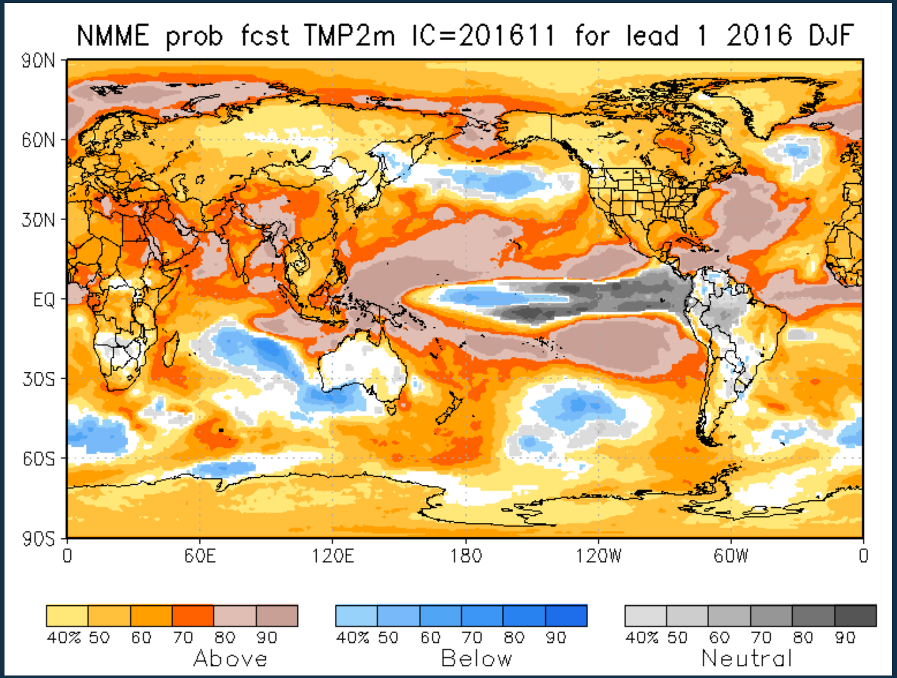
Representation of Teleconnections





Deterministic and Probabilistic forecasts

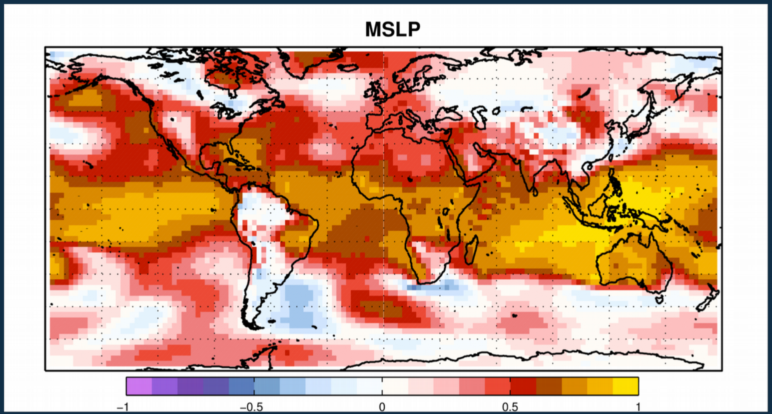


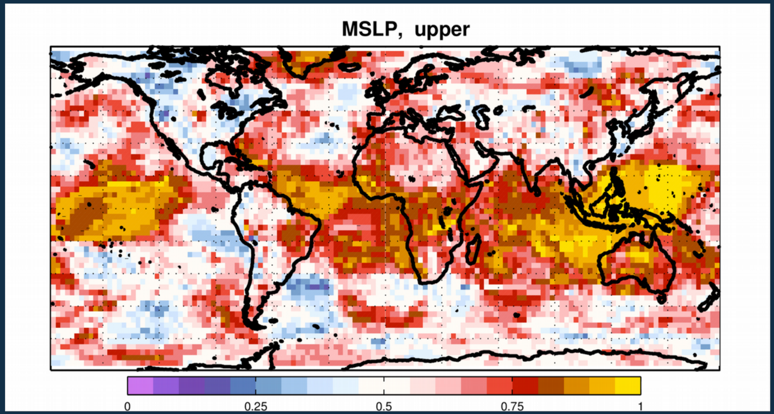


How to evaluate them?

ACC
Anomaly Correlation Coefficient

ROC
Relative Operative Characteristics Score





Deterministic

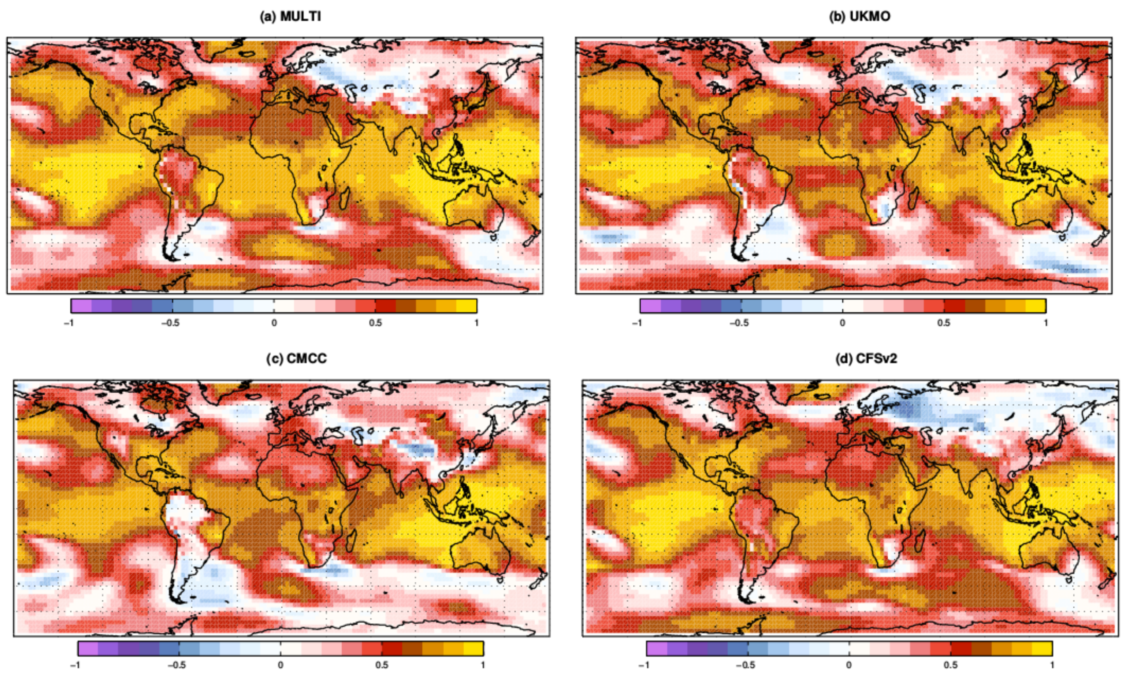
Probabilistic

Anomaly Correlation Coefficient

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

- *It is computed for the ensemble mean forecast*
- *It is not affected by bias correction / calibration*
- *Also used as pattern correlation (Taylor diagrams)*
- *Forecasts errors can be large regardless of ACC*
- *ACC increases with the ensemble size*
- *ACC deteriorates with lead time*

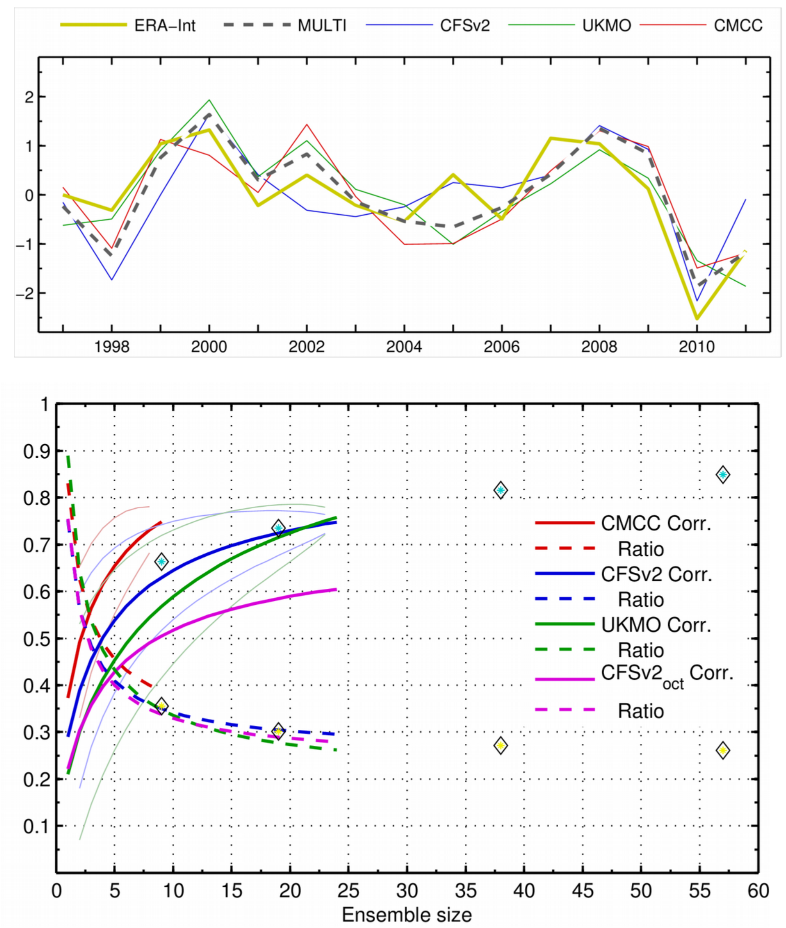
Common features indicate real predictability



Deterministic predictive skill (ACC) for the DJF mean MSLP ensemble mean anomalies of each SPS and the MULTI. Based on a one-sided T-test accounting for auto-correlation (see text), all correlations above 0.50 (dark red, brown and yellow shading) are statistically significant at least at the 0.95 level.

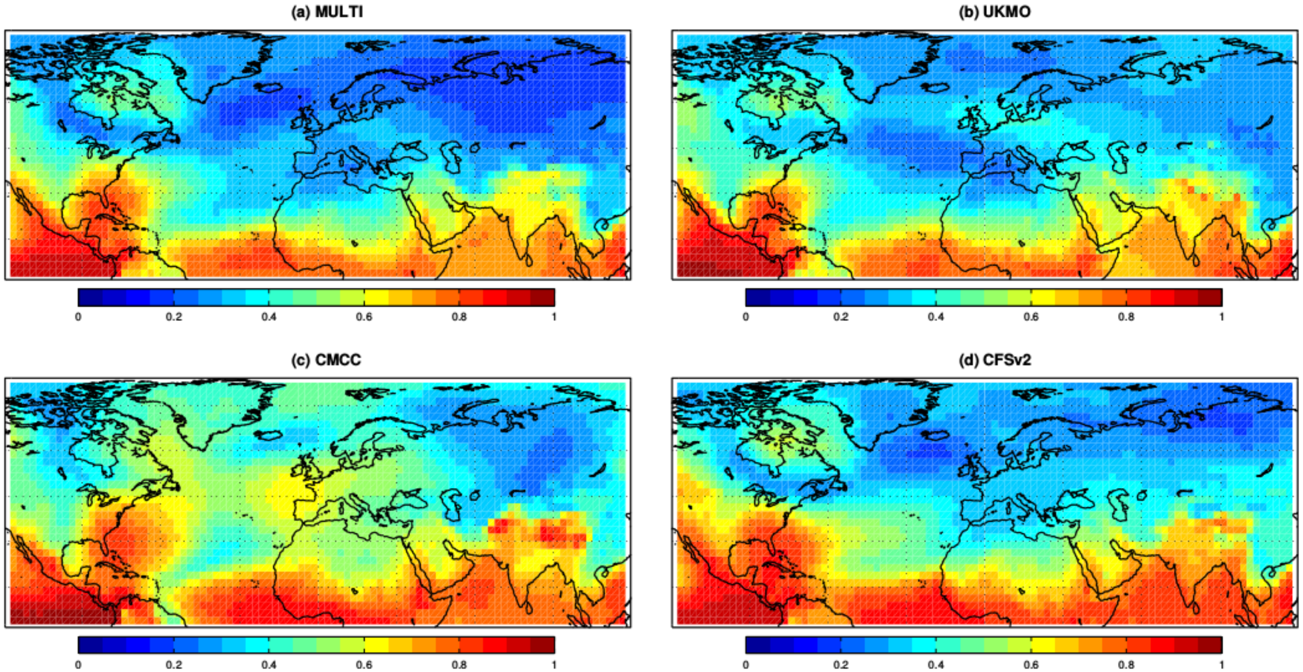
Athanasiadis et at., 2016

Example: the winter NAO



Athanasiadis et at., 2016

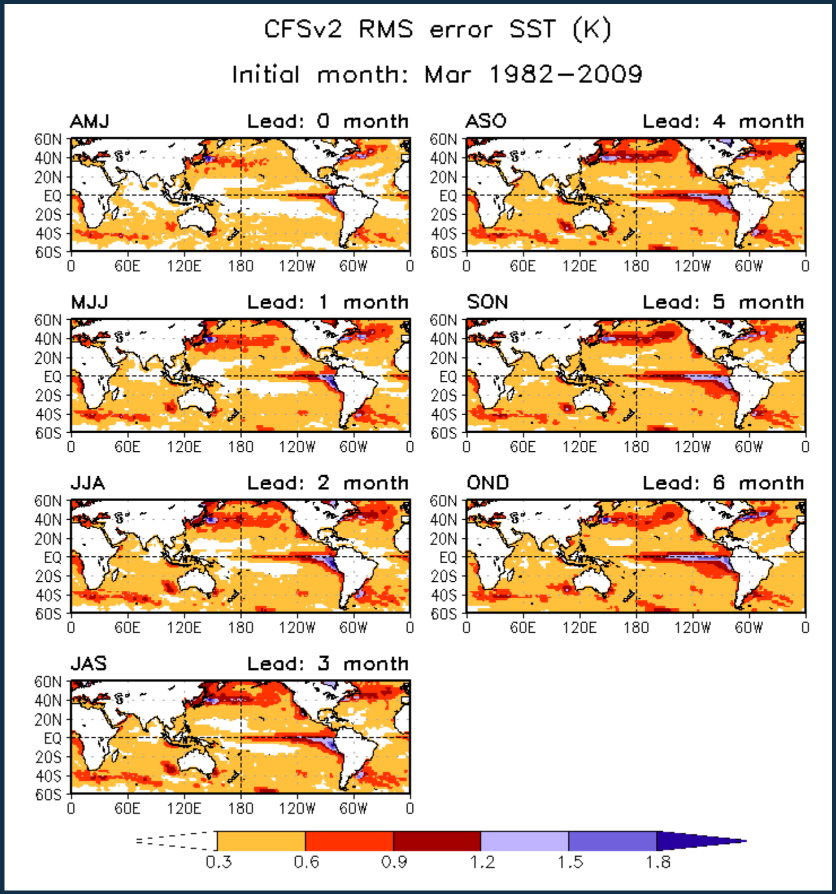
Signal-to-noise ratio

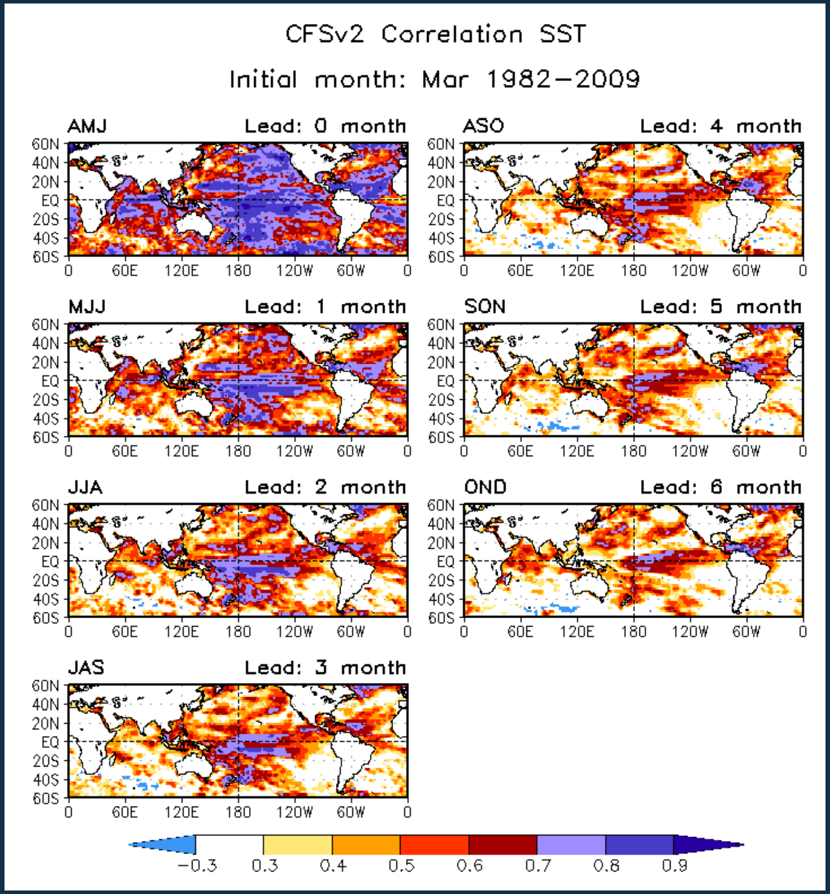


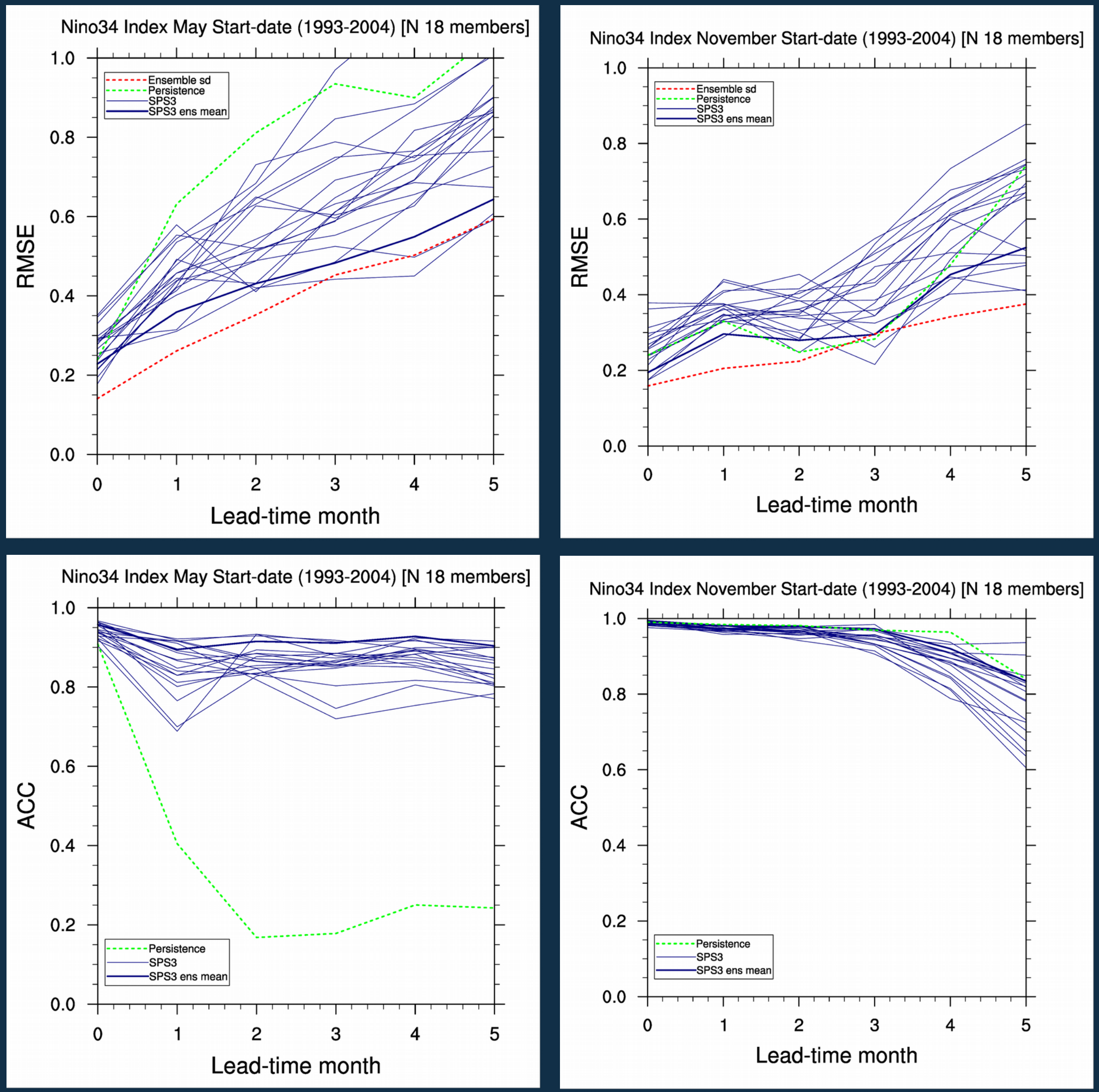
Athanasiadis et at., 2016

Signal-to-noise ratio for the DJF mean MSLP anomalies of each SPS and the MULTI.

RMSE and ACC depend on lead time







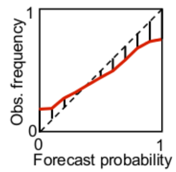
Nino34 Index May Start-date (1993-2004) [N 18 members]
RMSE
Lead-time month
Ensemble sd
Persistence
SPS3
SPS3 ens mean
Nino34 Index November Start-date (1993-2004) [N 18 members]
RMSE
Lead-time month
Nino34 Index May Start-date (1993-2004) [N 18 members]
ACC
Lead-time month
Nino34 Index November Start-date (1993-2004) [N 18 members]
ACC
Lead-time month

How to assess the probabilistic skill?
Reliability diagrams and ROC score / curves

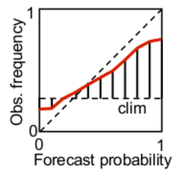
Reliability: Proximity to diagonal

Resolution: Variation about horizontal (climatology) line

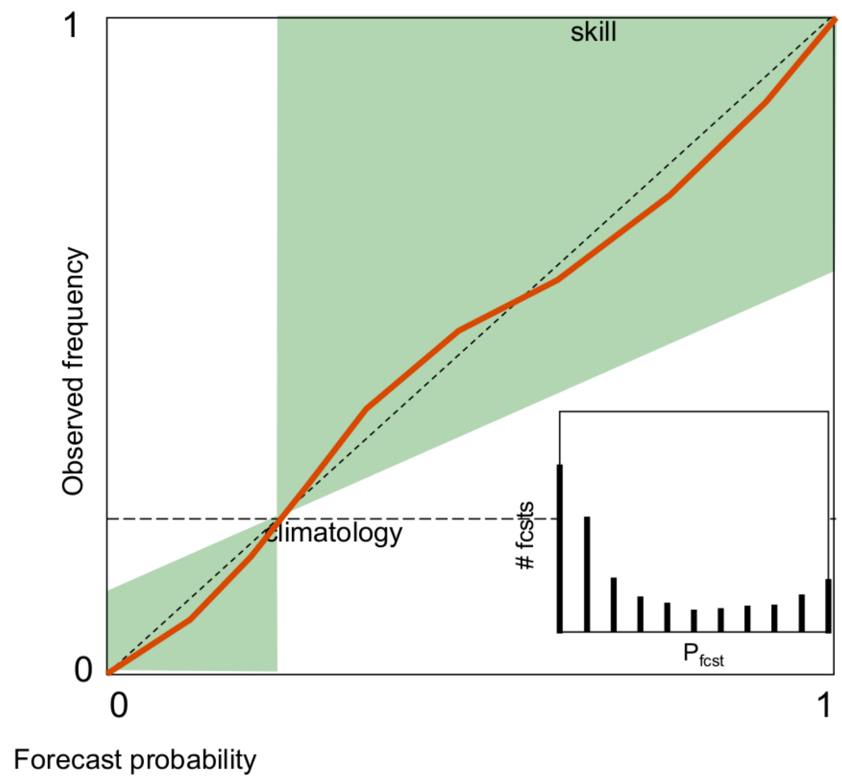
No skill line: Where reliability and resolution are equal – Brier skill score goes to 0

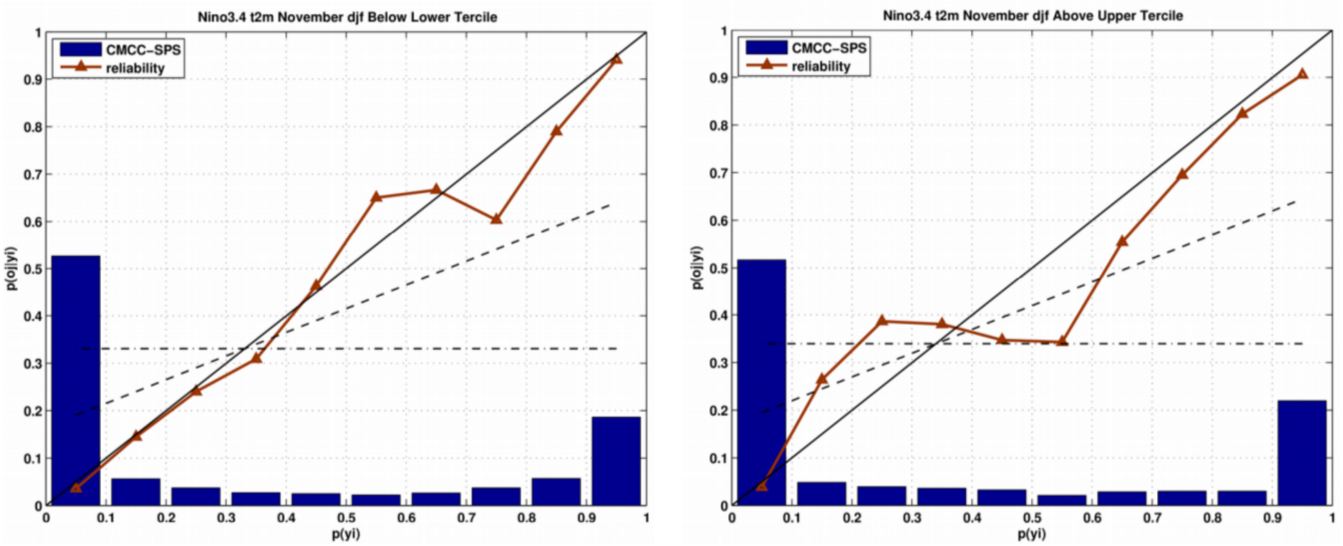


Reliability



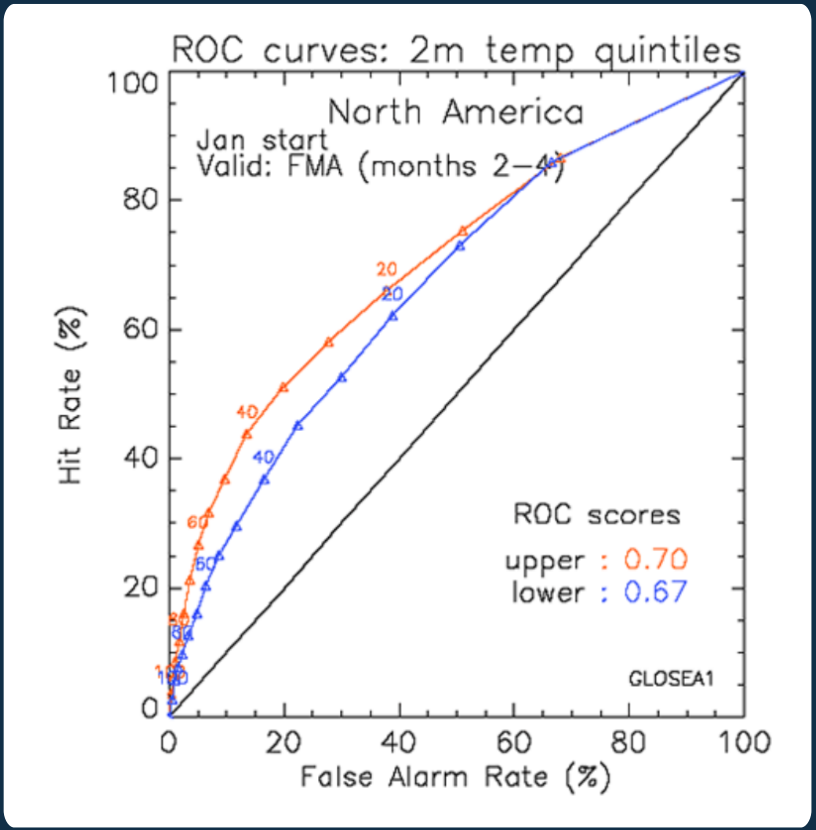
Resolution



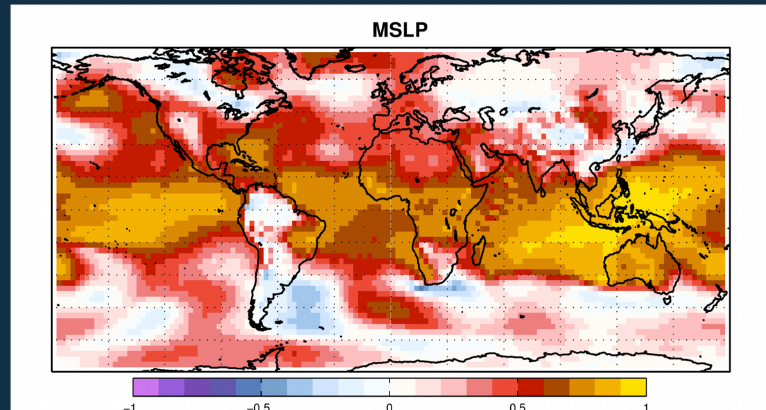


Reliability diagrams for T2M at the NINO3.4 area for the upper and lower third of the distribution as defined by the corresponding terciles. These are for DJF from the CMCC_SPS_v2 hindcasts initialized in November (1989–2005). *Materia et at., 2014*

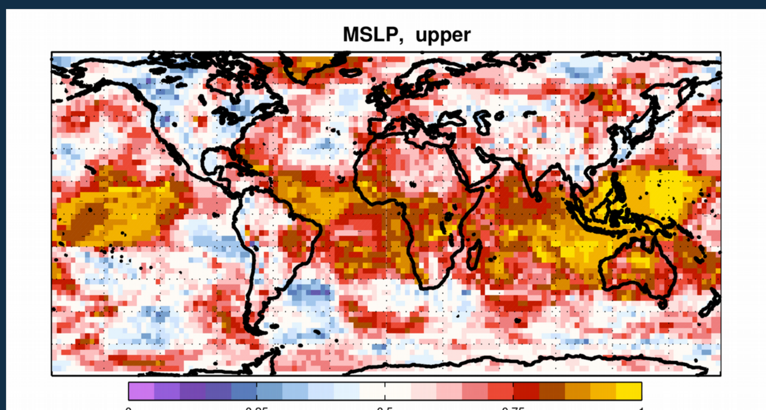
Relative Operative Characteristics (ROC) curve



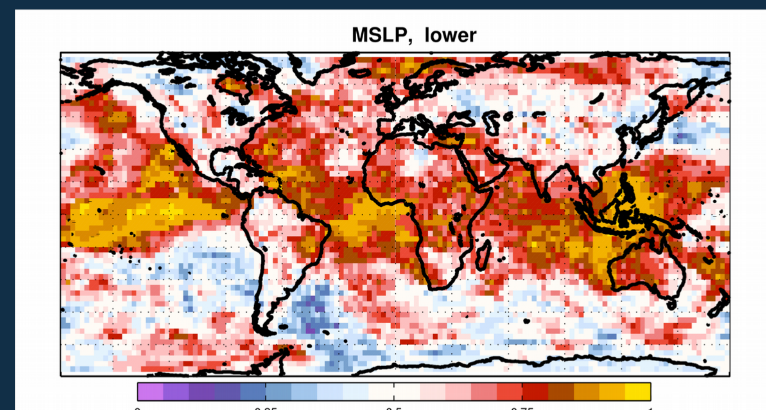
ACC

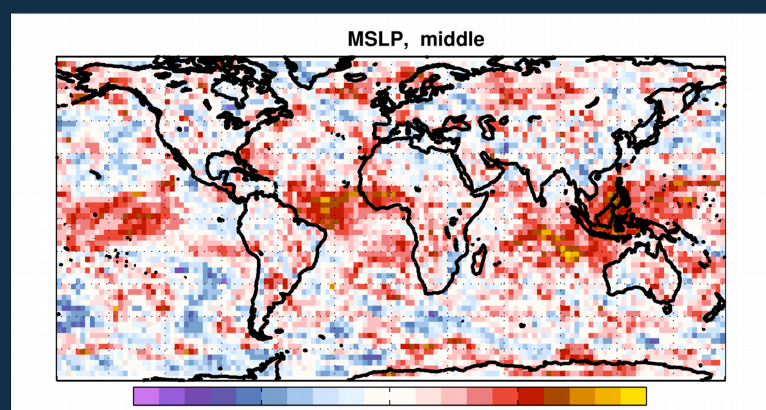

MSLP
−1
−0.5
0
0.5
1

ROC +

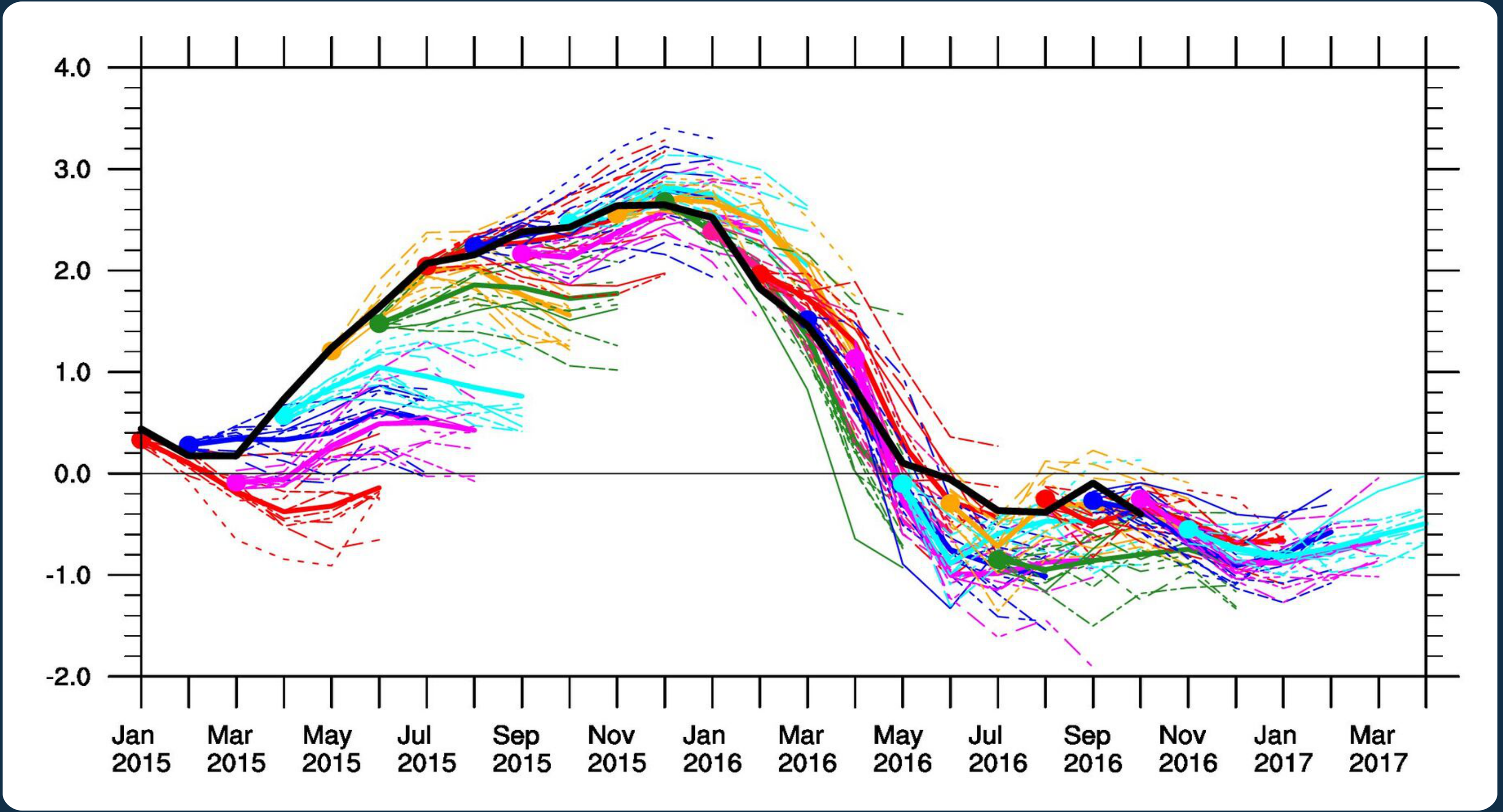

MSLP, upper
0
0.25
0.5
0.75
1

ROC -

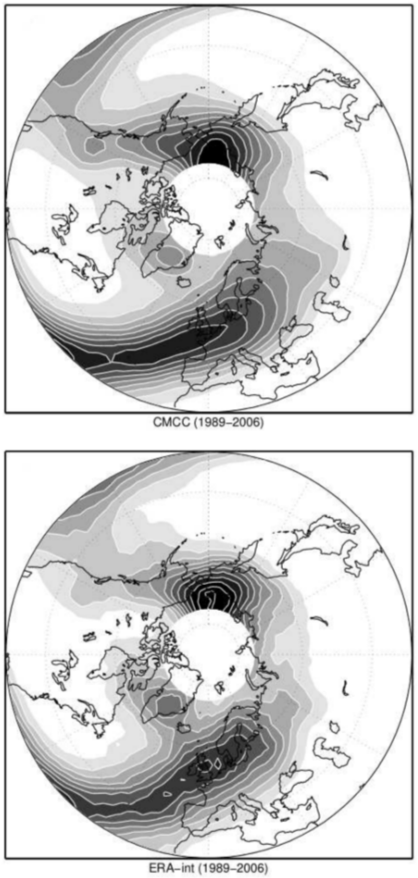

MSLP, lower
0
0.25
0.5
0.75
1

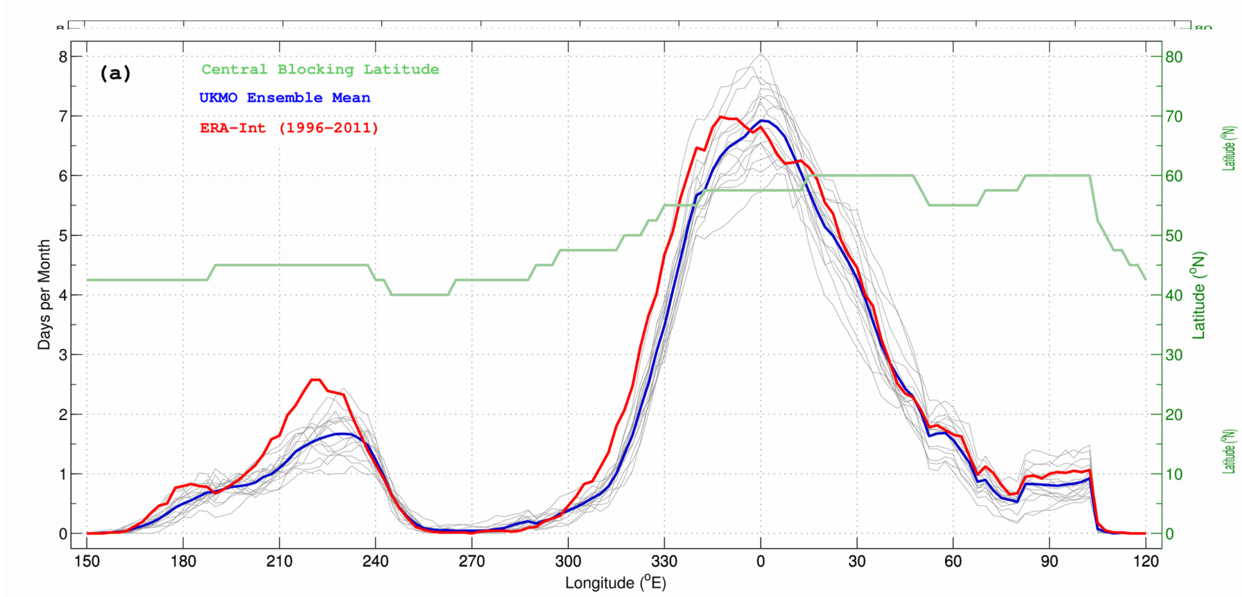

MSLP, middle
0
0.25
0.5
0.75
1

Ensemble plumes for NINO3.4



Representation of blocking frequency





Blocking frequency in DJF for CMCC_SPS_v1.5 and the ERA-Interim. Instantaneous blocking on the left, persistent large-scale blocking on the right. The contours show % of blocked days, contour intervals are 2 and 0.5, correspondingly. See main text for further details.

Athanasiadis et at., 2014

Taylor diagrams to assess spatial match

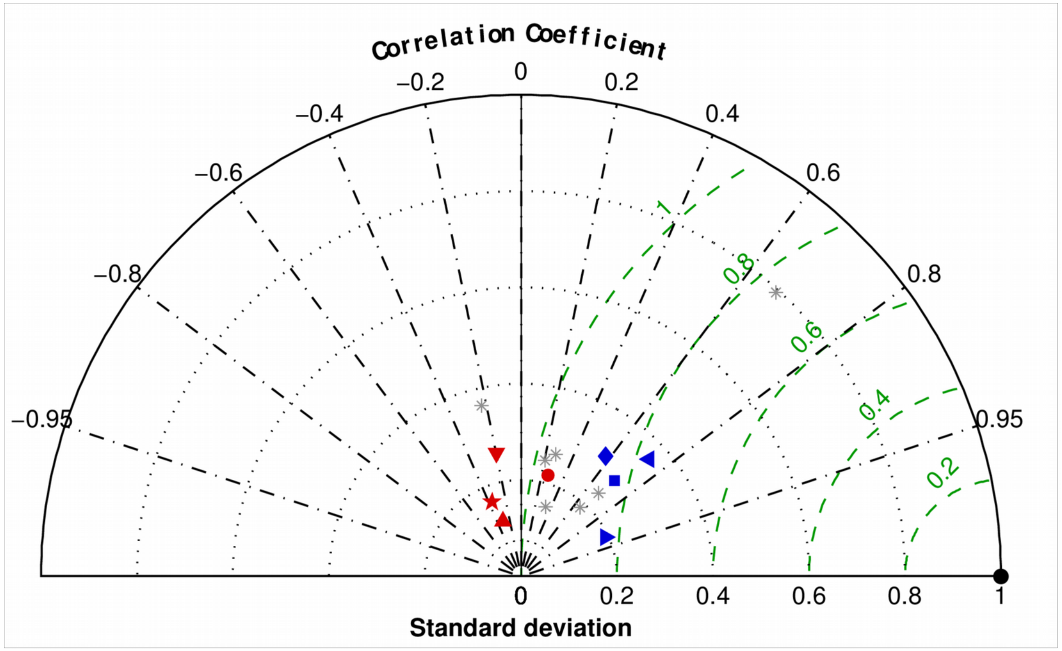
$$R = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(f_n - \bar{f}\right)\left(r_n - \bar{r}\right)}{\sigma_f \sigma_r}$$

$$E'^2 = \frac{1}{N}\sum_{n=1}^{N}\left[\left(f_n - \bar{f}\right) - \left(r_n - \bar{r}\right)\right]^2$$

$$\sigma_f^2 = \frac{1}{N}\sum_{n=1}^{N}\left(f_n - \bar{f}\right)^2$$

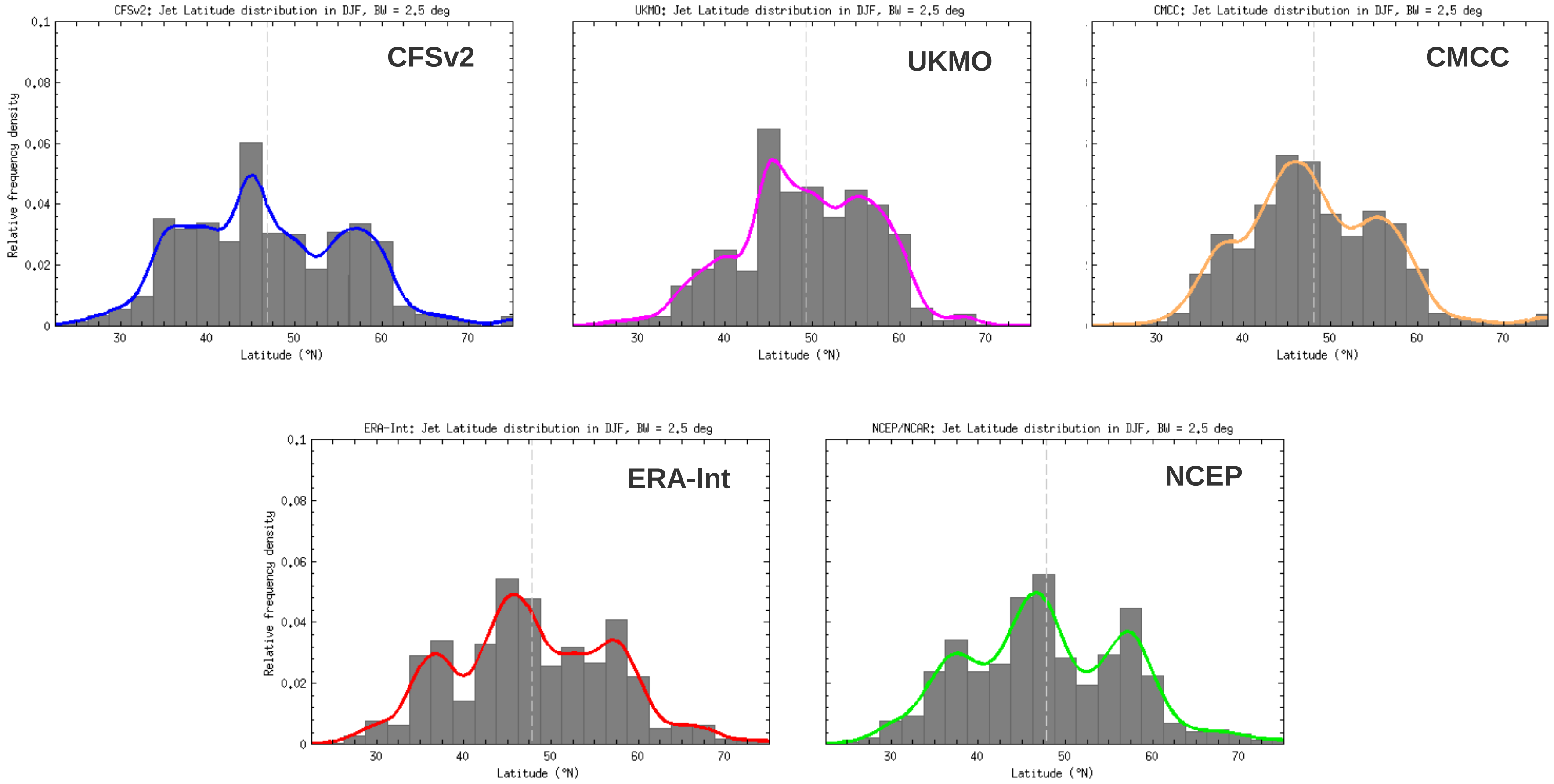
$$\sigma_r^2 = \frac{1}{N}\sum_{n=1}^{N}\left(r_n - \bar{r}\right)^2$$

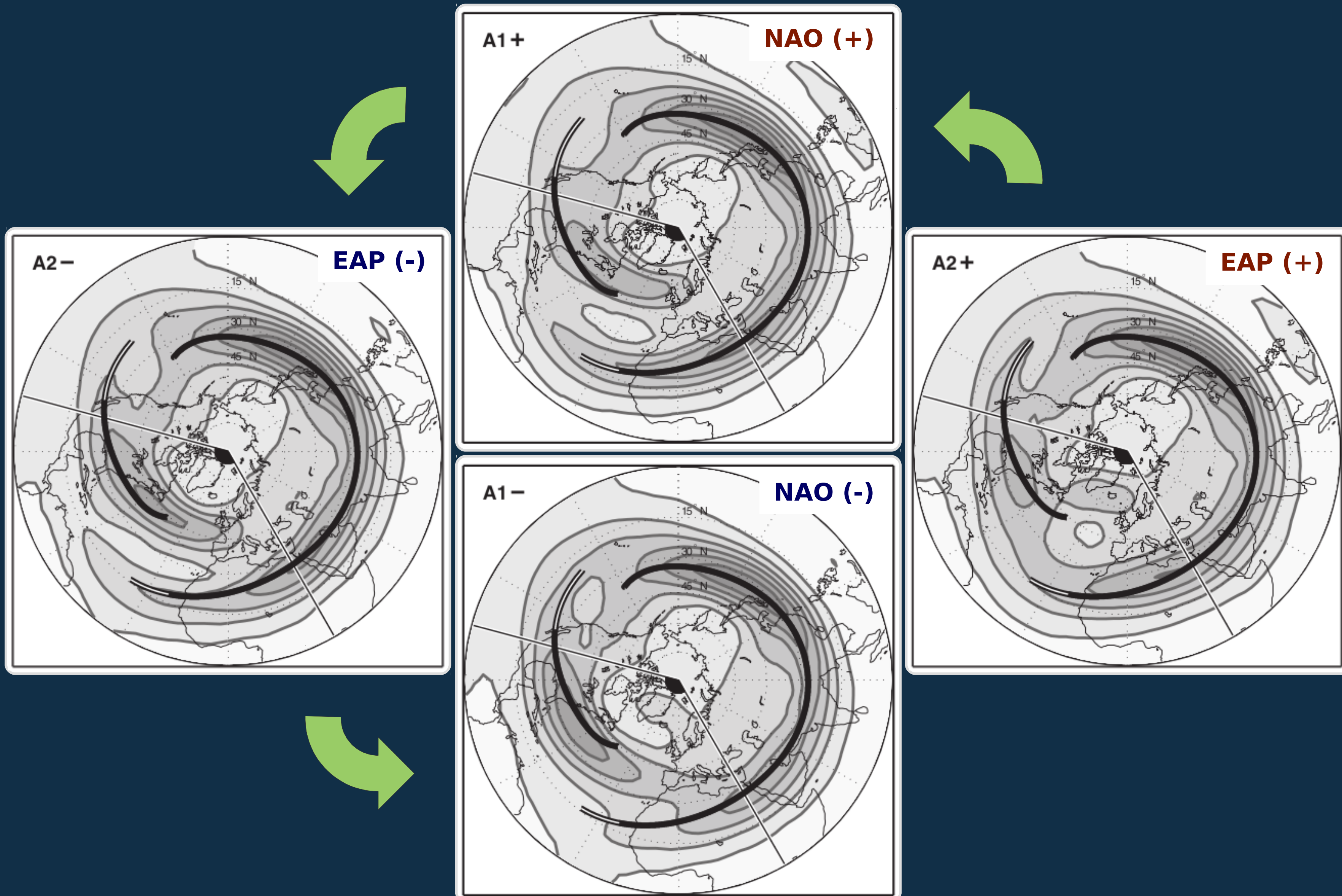
$$E^2 = \bar{E}^2 + E'^2$$



Taylor K.E. (2001)

For middlatitude dynamics and variability the eddy-driven jet variability is key





P. Athanasiadis, M. Wallace and J. Wettstein, J. Atm. Sci., (2010)

SKILL COMES LAST,
FIRST COMES MODEL PERFORMANCE

- *In order for a seasonal prediction system to make skillful forecasts, it has to simulate well all the major physical processes associated to seasonal predictability.*
- *This is possible only if the model has a realistic climate and variability thereof.*
- *Variability and predictability at seasonal timescales are directly associated to teleconnections and the drivers of low-frequency variability.*
- *A large ensemble size and a sophisticated initialization strategy are necessary for sampling the non-predictable, chaotic variability and enhancing the signal-to-noise ratio.*
- *Deterministic and probabilistic skills provide complementary information.*