“New” errors in Nino 3.4 forecasts

Michael Tippett
Columbia University

Michelle L’Heureux, Arun Kumar, and Emily Becker
NOAA/NWS/NCEP/Climate Prediction Center

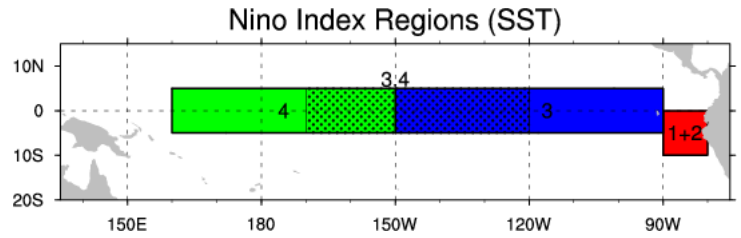
Summary

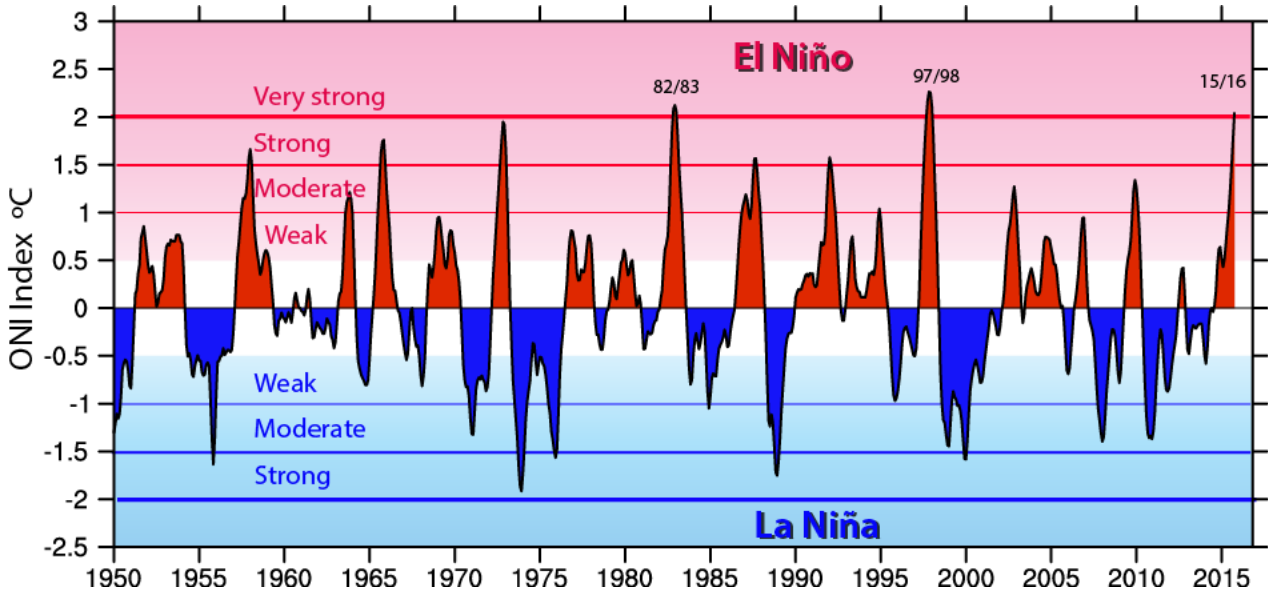
- Late-spring ENSO *forecasts errors* depend on the observed ENSO tendency before the forecast.
  - “Momentum” = ENSO tends to keep going in the same direction
  - Models have state-dependent errors
- Consequence: Late-spring ENSO forecasts are unreliable.
- Are model dynamics or initialization to blame?
  - Some evidence for dynamics being the culprit
- EOF analysis shows ENSO time structure is 2-dimensional
  - Models with better time structure have smaller errors.
  - EOF-based regression reduces forecast MSE

https://www.climate.gov/news-features/blogs/enso/what-el-niño–southern-oscillation-enso-nutshell

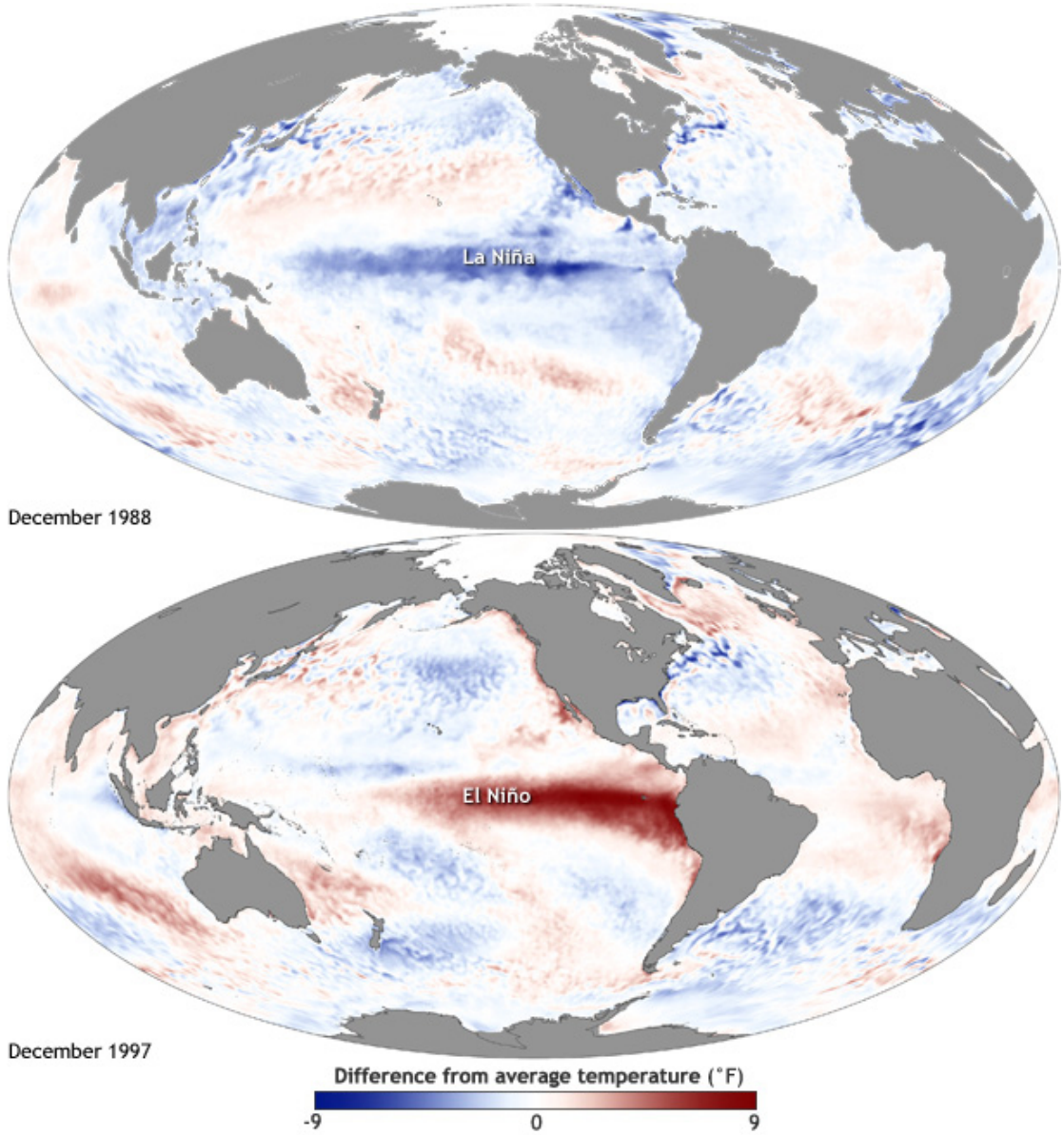
What is ENSO?

- El Nino, La Nina, and Nino 3.4*





https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni

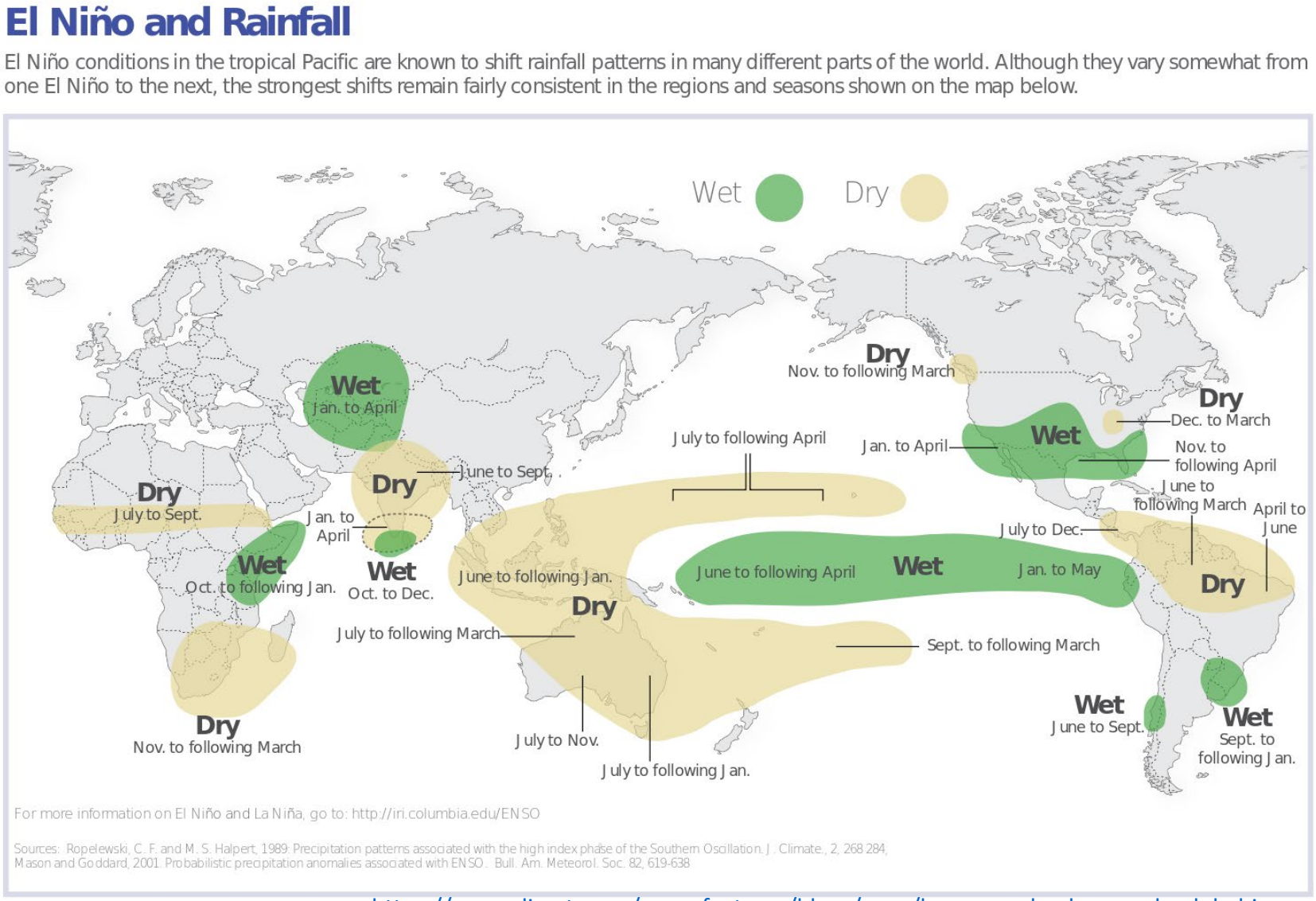


*No flavors today

Why is ENSO is important?

Predictable impacts

Temperature, precipitation, hurricanes, tornadoes, human conflict …



"El Niño's effects on Europe are controversial, complex and difficult to analyze, as it is one of several factors that influence the weather over the continent and other factors can overwhelm the signal." Wikipeida

https://www.climate.gov/news-features/blogs/enso/how-enso-leads-cascade-global-impacts

How is ENSO predicted?

L'Heureux, M., A. Levine, M. Newman, C. Ganter, J.-J. Luo, M. Tippett, and T. Stockdale, 2019: "ENSO Prediction". In El Niño-Southern Oscillation (ENSO) in a Changing Climate, M. McPhaden, A. Santoso, W. Cai, eds, in press.

Main forecast guidance:

- Initialized coupled atmosphere-ocean models
  - Systematic errors (biases) are removed

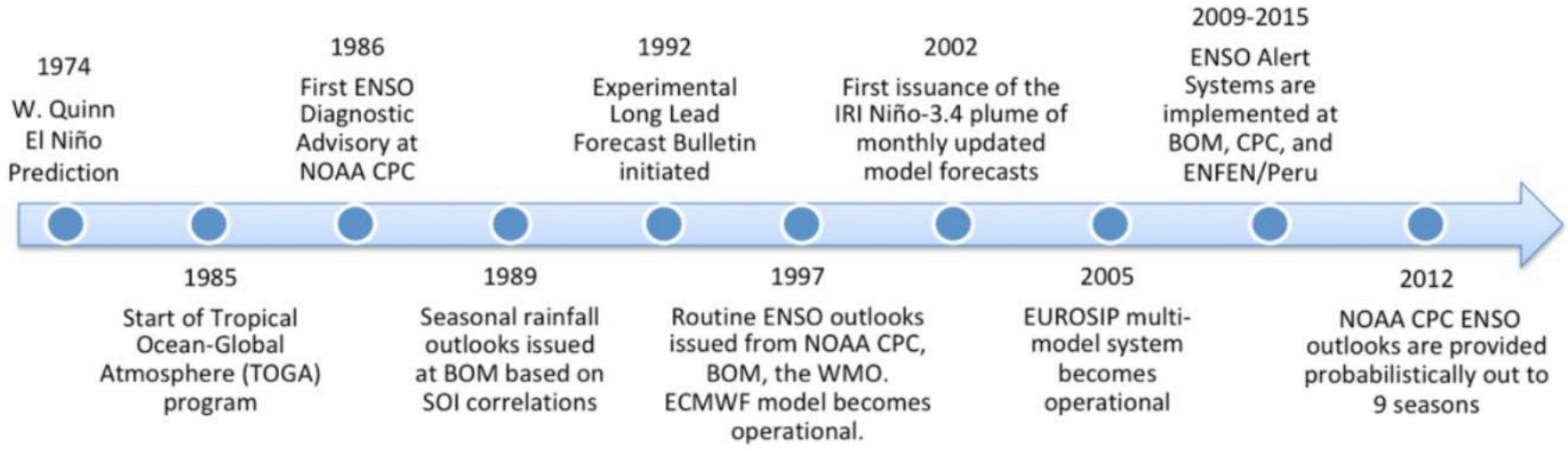
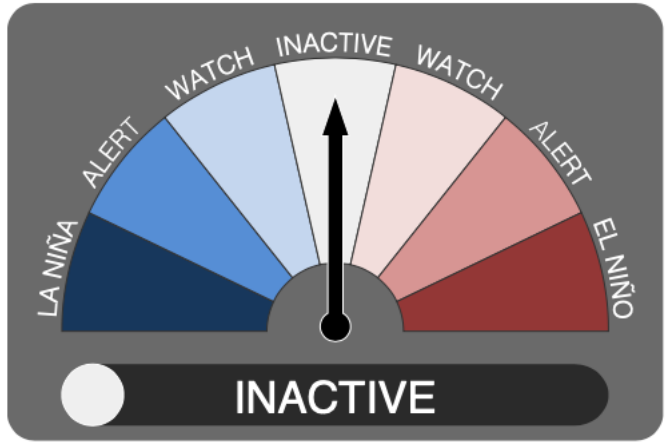


Figure 1. A timeline of events in ENSO forecasting.

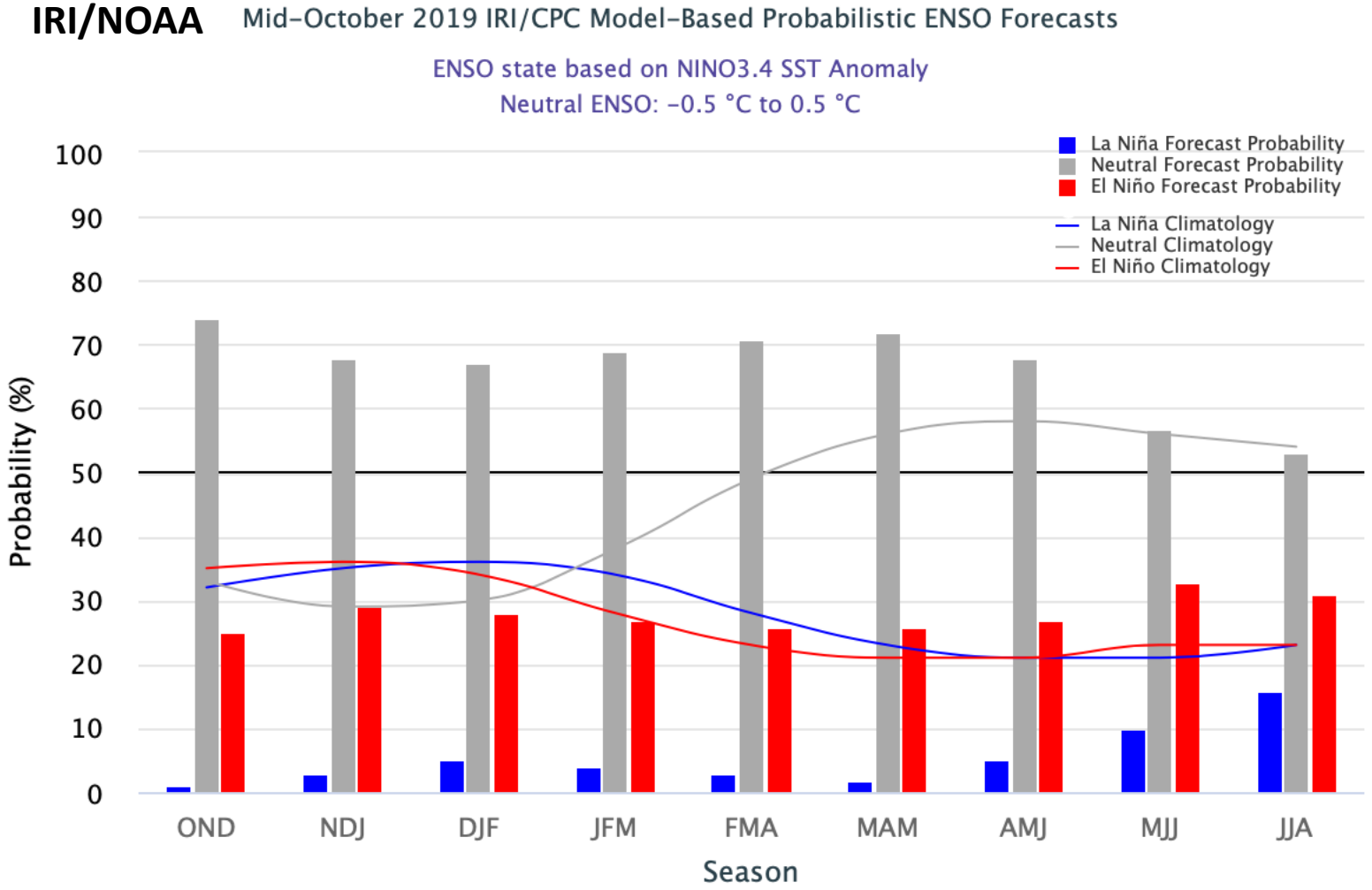
Categorical

What do ENSO forecasts look like?

BoM



IRI/NOAA



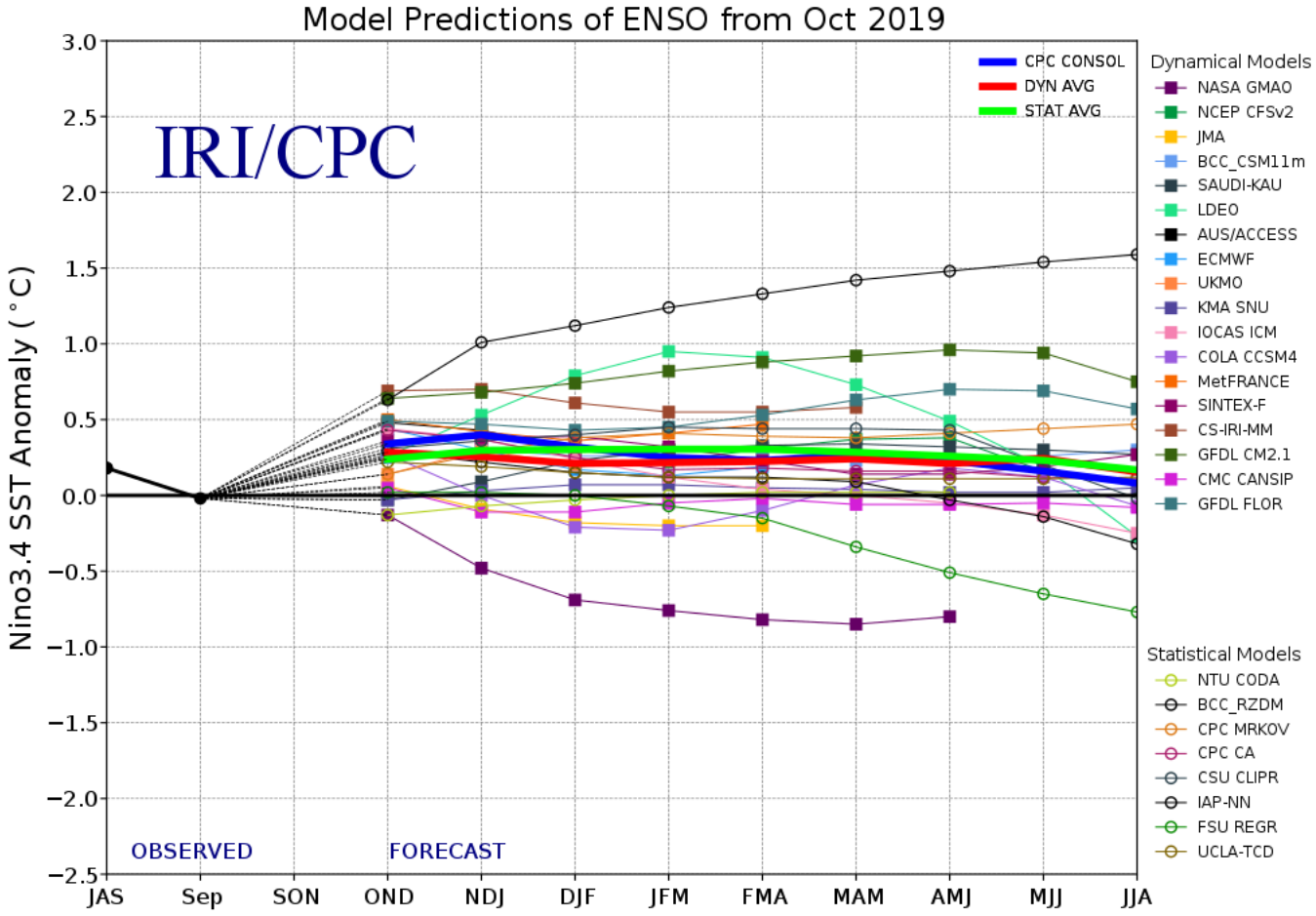
JMA

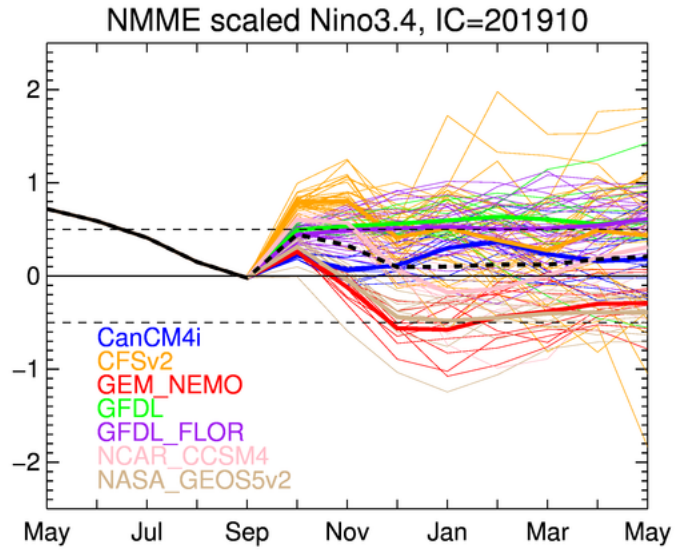
YEAR	MONTH	mean period	El Niño	ENSO neutral	La Niña
2019	AUG	JUN2019–OCT2019		100	
	SEP	JUL2019–NOV2019		100	
	OCT	AUG2019–DEC2019		90	10
	NOV	SEP2019–JAN2020	10	70	20
	DEC	OCT2019–FEB2020	10	70	20
2020	JAN	NOV2019–MAR2020	20	60	20
	FEB	DEC2019–APR2020	20	60	20

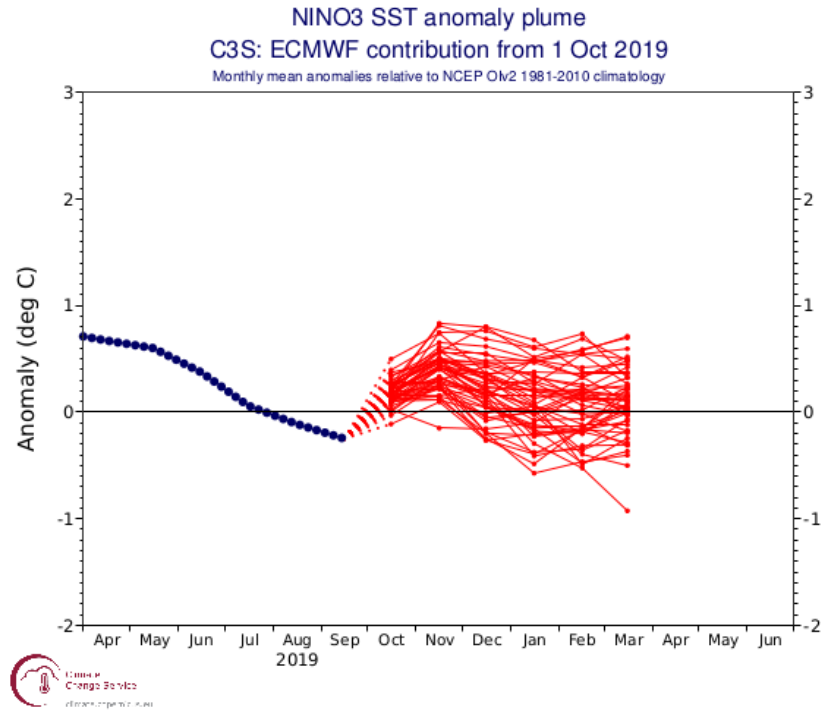
Continuous

NMME

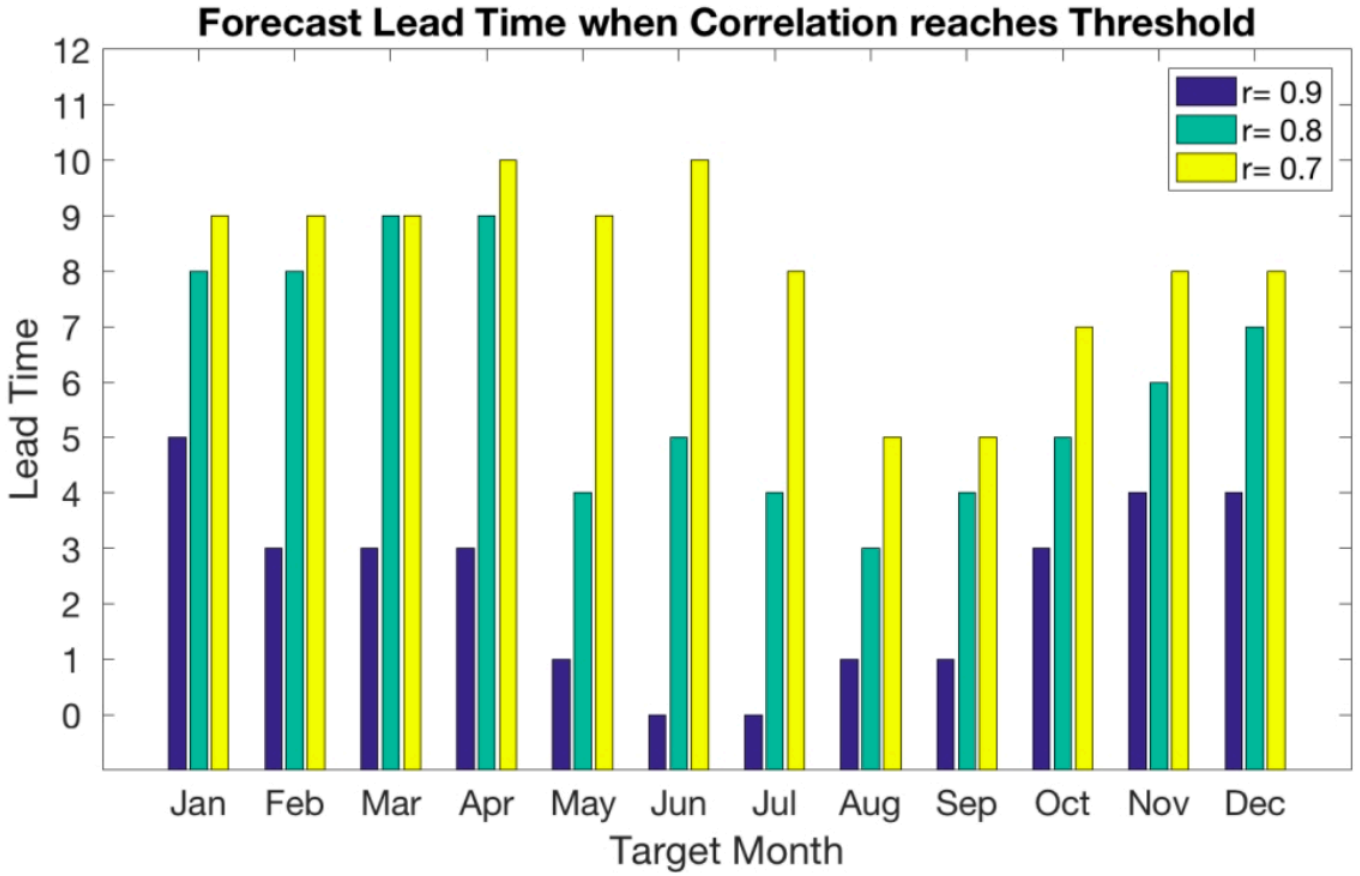
What do ENSO forecasts look like?







How skillful are ENSO predictions?

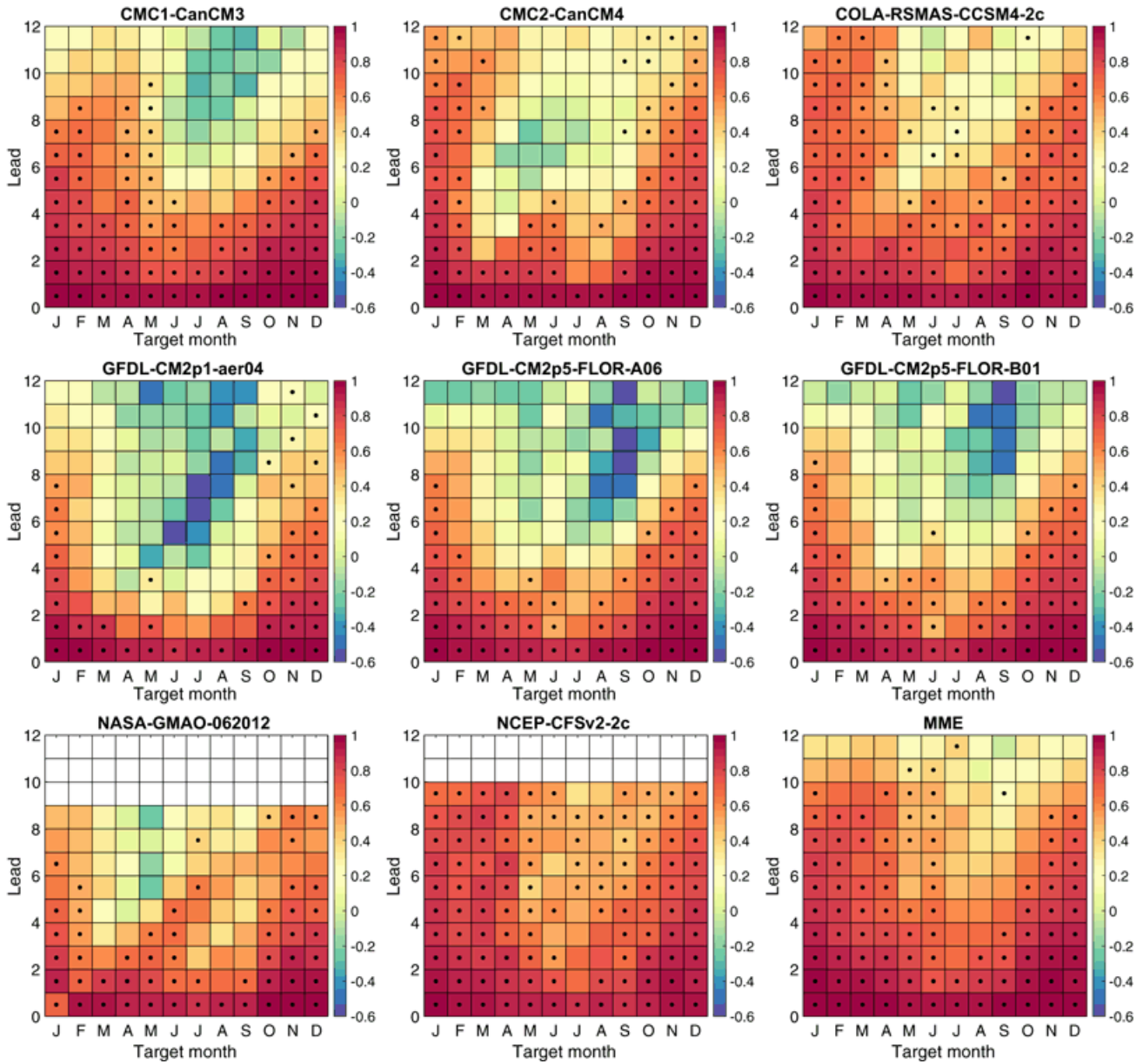


L'Heureux, M., A. Levine, M. Newman, C. Ganter, J.-J. Luo, M. Tippett, and T. Stockdale, 2019: "ENSO Prediction". In El Niño-Southern Oscillation (ENSO) in a Changing Climate, M. McPhaden, A. Santoso, W. Cai, eds, in press.

Figure 3. The forecast lead-time (y-axis) when Niño-3.4 index skill exceeds certain correlation thresholds. On the x-axis, the target month of the forecast is presented. Model data is based on the ensemble average of ~100 members from the North American Multi-Model Ensemble from 1982 to 2018. Departures in the Niño-3.4 index are based on 1982-2010 monthly averages.

"The spring barrier causes forecasts that traverse the months of April through June to tend to have less favorable verification than those that do not go through those months."

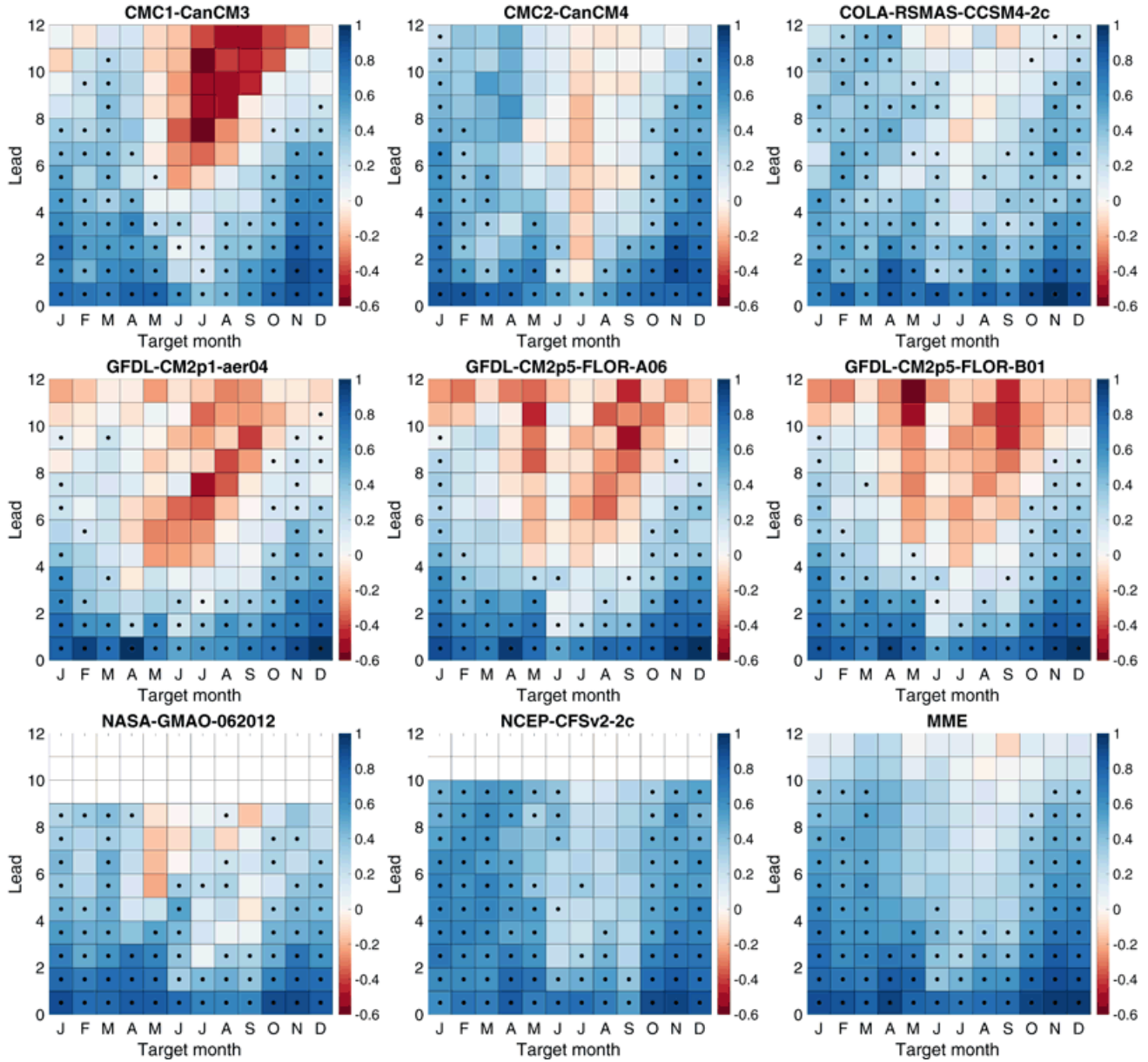
Mean-squared error skill score for the NMME



A. G. Barnston, M. K. Tippett, M. Ranganathan, and M. L. L'Heureux. Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. Clim. Dyn., 2017.

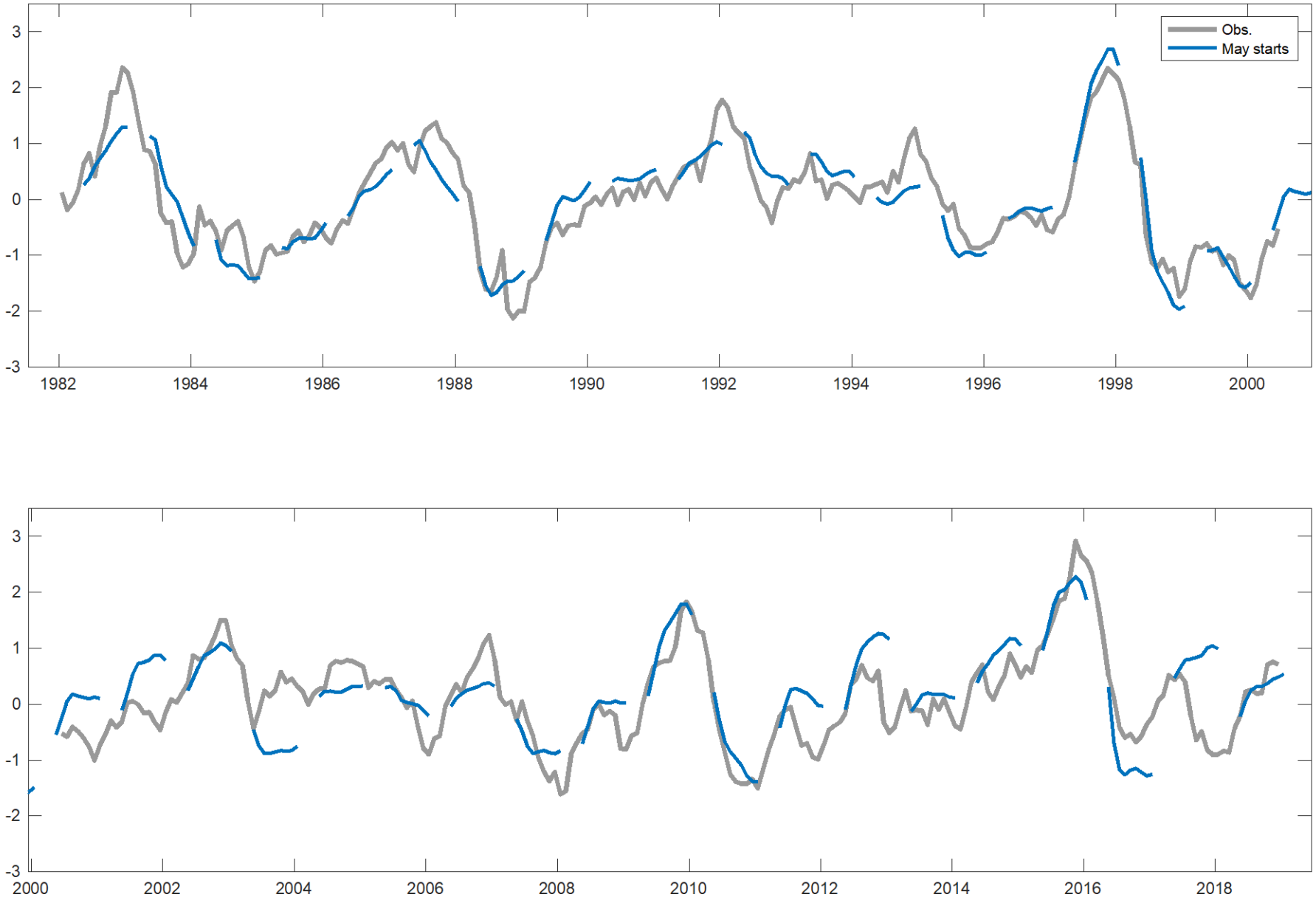
"The spring barrier causes forecasts that traverse the months of April through June to tend to have less favorable verification than those that do not go through those months."

Rank probability skill score for the NMME

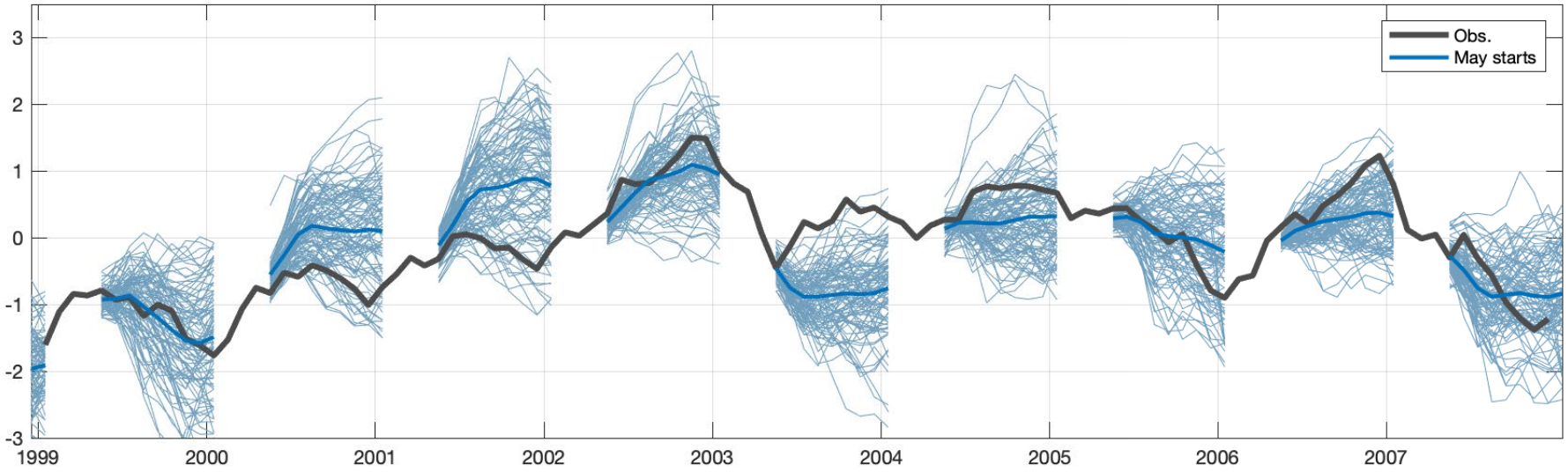


M. K. Tippett, M. Ranganathan, M. L'Heureux, A. G. Barnston, and T. DelSole. Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. Clim. Dyn., 2017

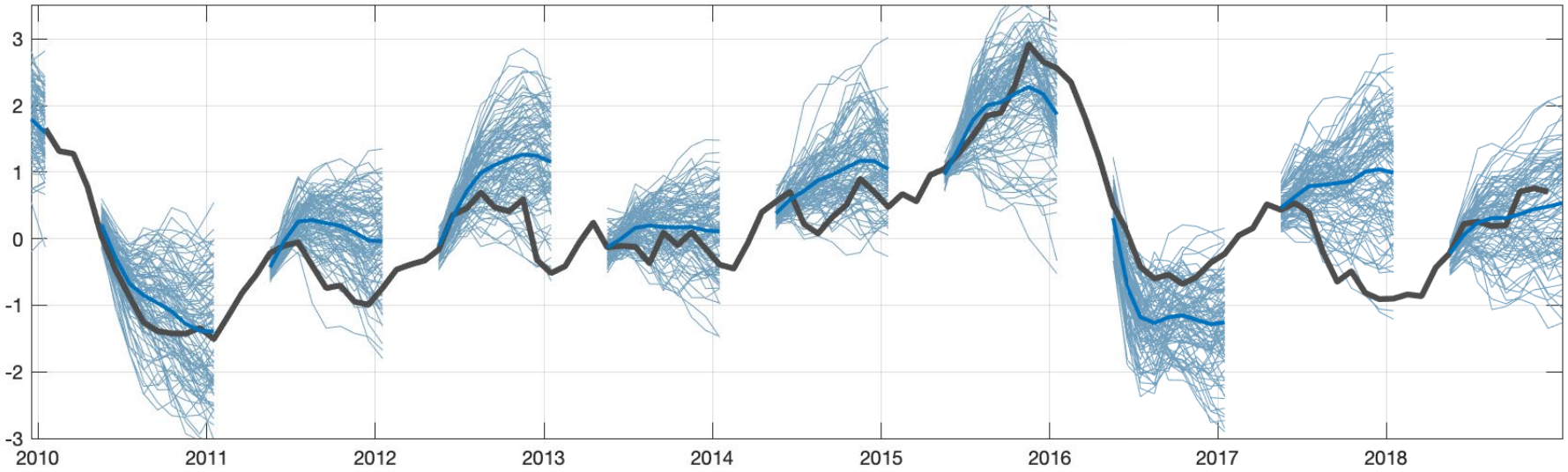
Nino 3.4 and May NMME ensemble-mean forecasts 1982—2018



Nino 3.4 and May NMME ensemble forecasts 1999—2018



Doesn't look great when the observed and forecast trajectories part ways



100 members

Is something wrong with these forecasts?

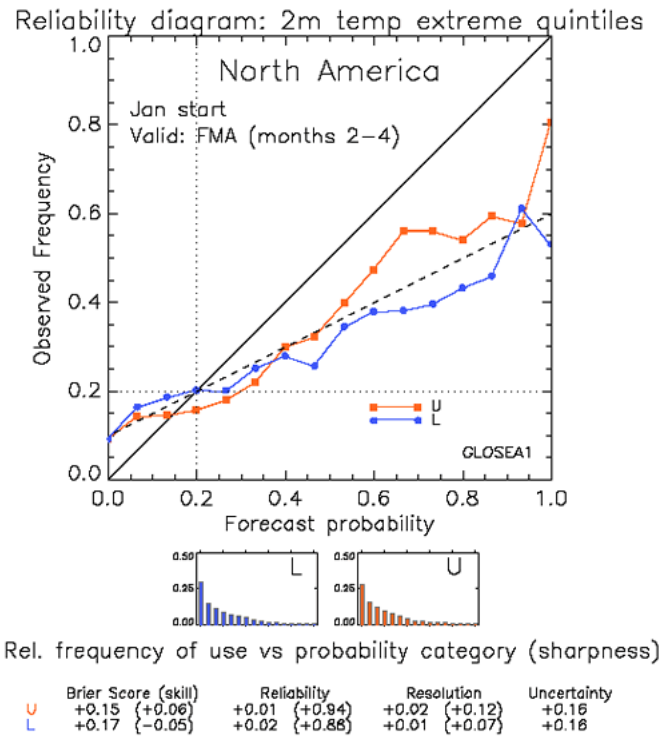
- Skill is low for spring forecasts.
  - Is this what low skill looks like?
- What does it mean for a probabilistic forecast to be wrong?
- Wrong = uncalibrated
  - Unreliable probabilities
  - Forecast ensemble that is is:
    - Over-dispersed (under-confident)
    - Under-dispersed (over-confident)
    - Biased

Focus on forecast tendencies

- Define the ***forecast tendency*** as lead-4 minus lead-1
  - For a Apr start, forecast tendency = Jul forecast – Apr forecast
  - For a May start, forecast tendency = Aug forecast – May forecast
- Compare forecast tendencies with observed ones
- Two forecast quantities
  - Sign of the tendency
    - “The probability of the tendency being positive is 90%”
  - Value of the tendency
    - The forecast distribution (pdf) of tendency values
- Are forecast tendencies reliable?

What does it mean for forecast *probabilities* to be reliable?

- (We have to look at many forecasts)
- When I forecast that the probability of an event (e.g., positive tendency) is X%, it occurs X% of the time.
- E[O | F] = F
  - O is 1 when the event occurs and 0 otherwise
  - F is the forecast probability
  - E means expectation (average)
- The graph of E[O | F] vs. F is called a reliability diagram

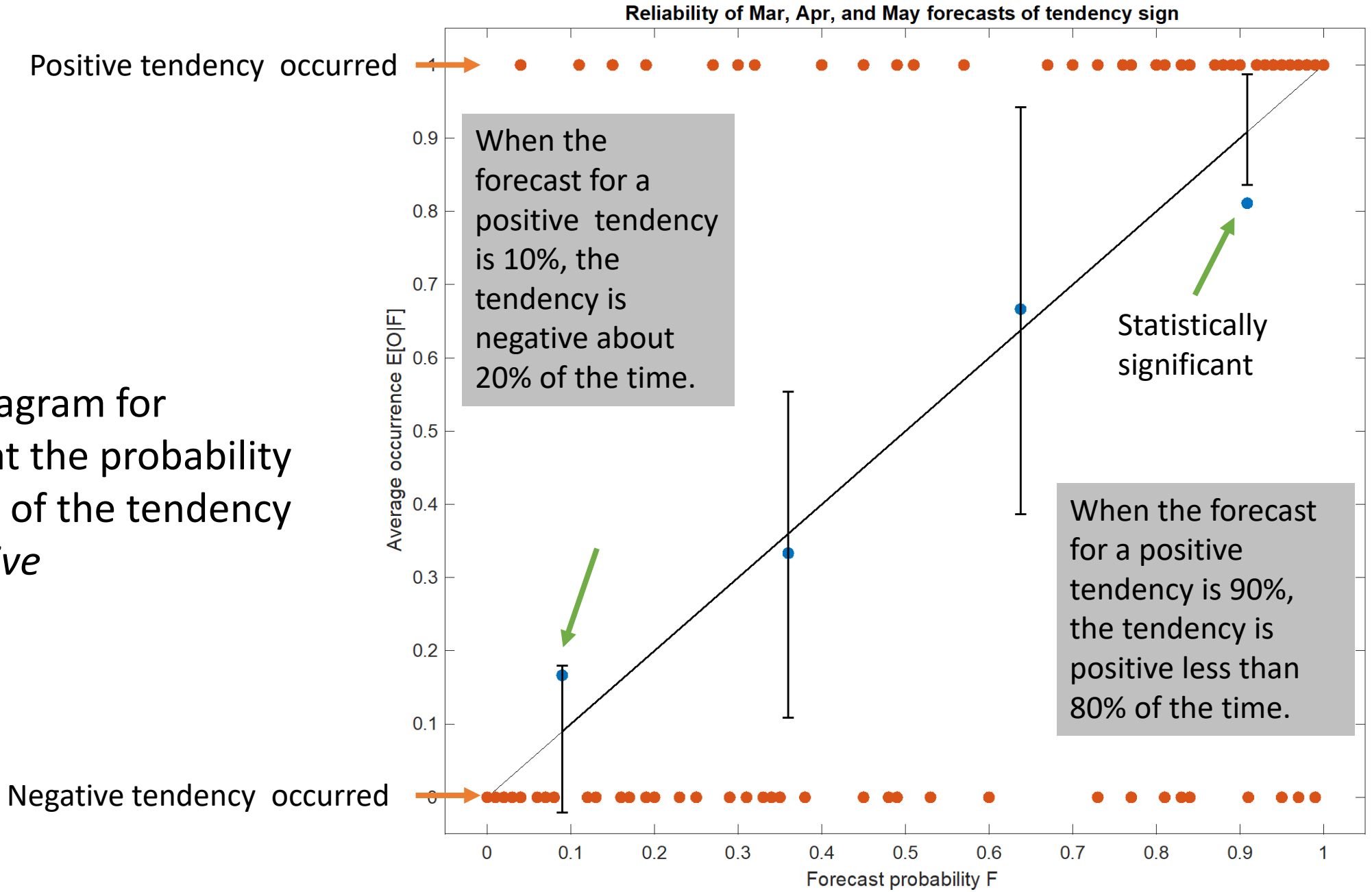


https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/interpret-reliability

Mar, Apr, May NMME forecasts of the sign of the ENSO tendency

Reliability diagram for forecasts that the probability that the sign of the tendency will be *positive*

Reliable?

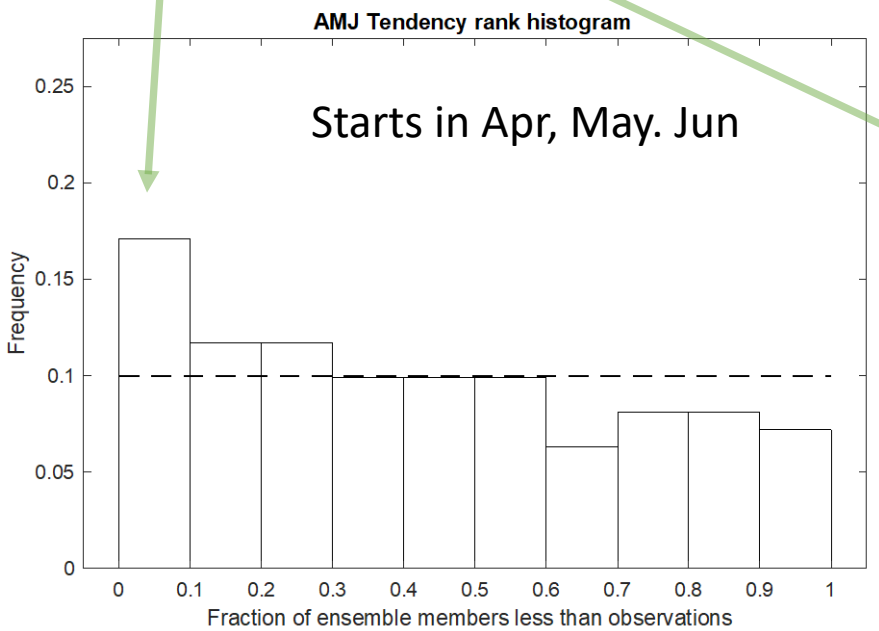


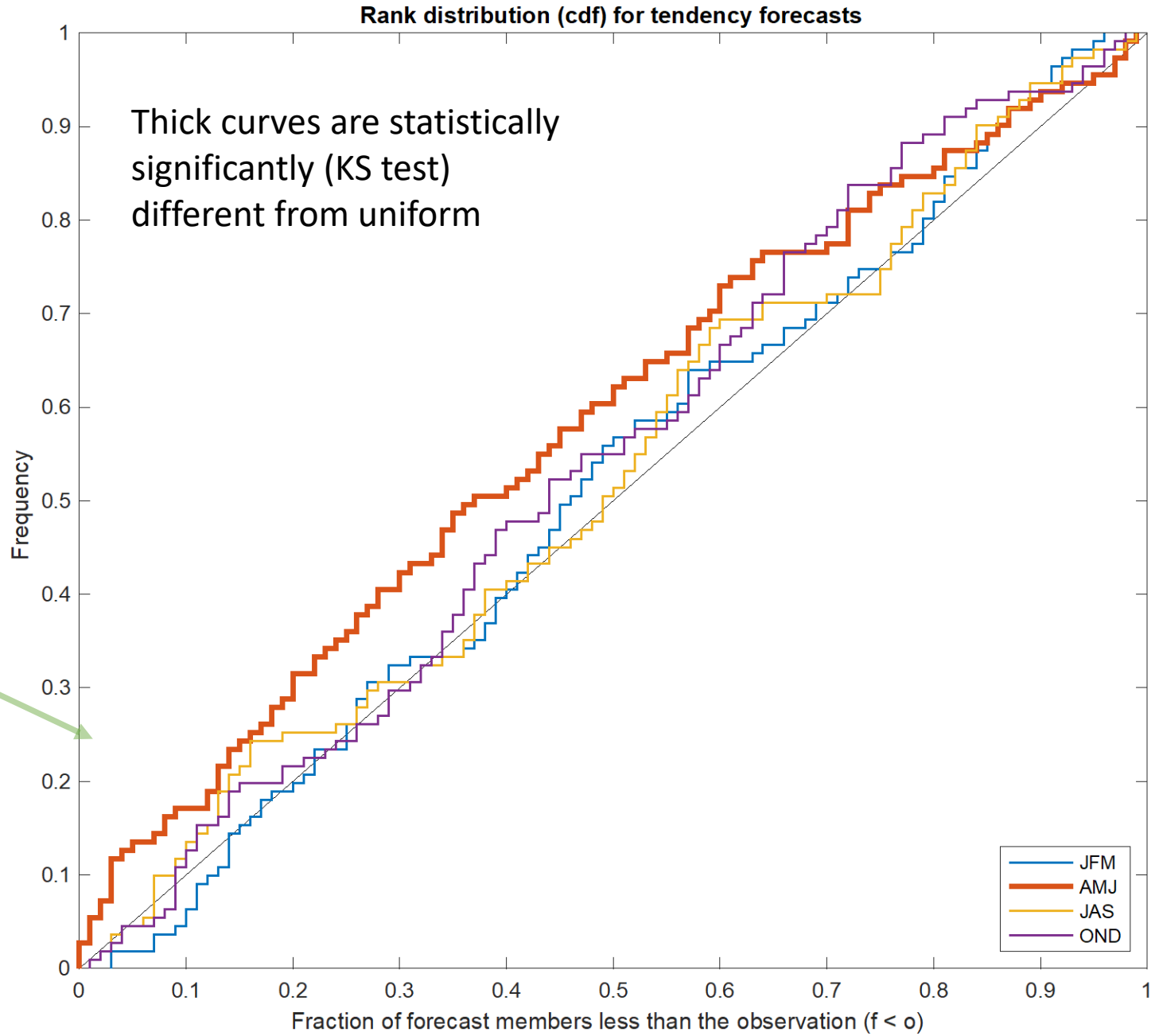
What does it mean for a forecast *ensemble* to be reliable?

- (We have to look at many forecasts)
- The observation (tendency value) is indistinguishable from a forecast ensemble member
  - The climatological distribution is a no-skill reliable forecast
- Rank histograms
  - The rank of the observations is uniformly distributed
    - If the observation value is less than all the ensemble values, its rank is 1.
    - If the observation value is greater than all the ensemble values, its rank is 101.
    - If the observation value is in the middle of the ensemble values, its rank is 51.
  - Divide rank by ensemble size = fraction of ensemble members less than observed
- On average, X% of the members should be less than the observations X% of the time.

When the ensemble lies mostly above the observation, the fraction of ensemble members less than the observation is small.

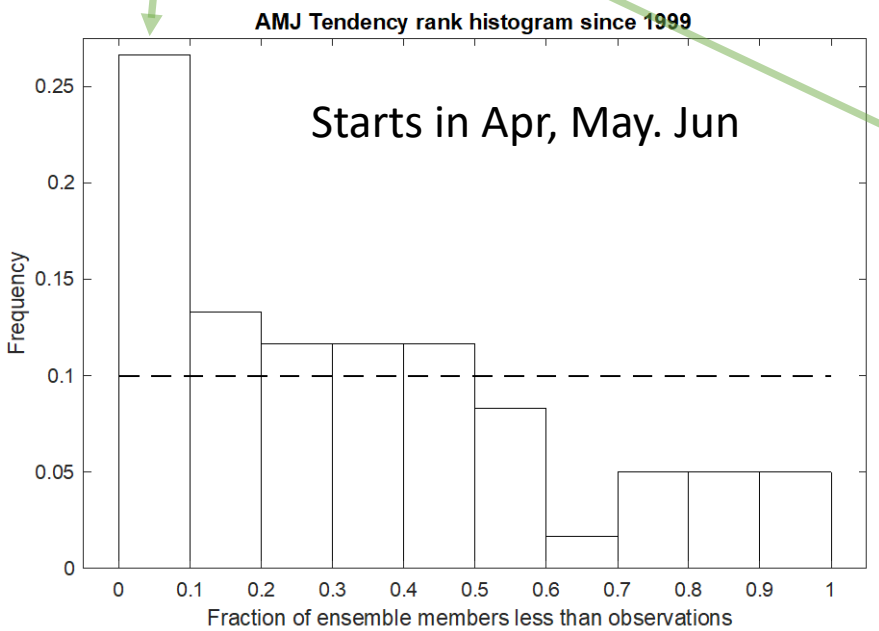
The fraction of ensemble members less than the observation is small too often, which means the ensemble lies above the observation too often

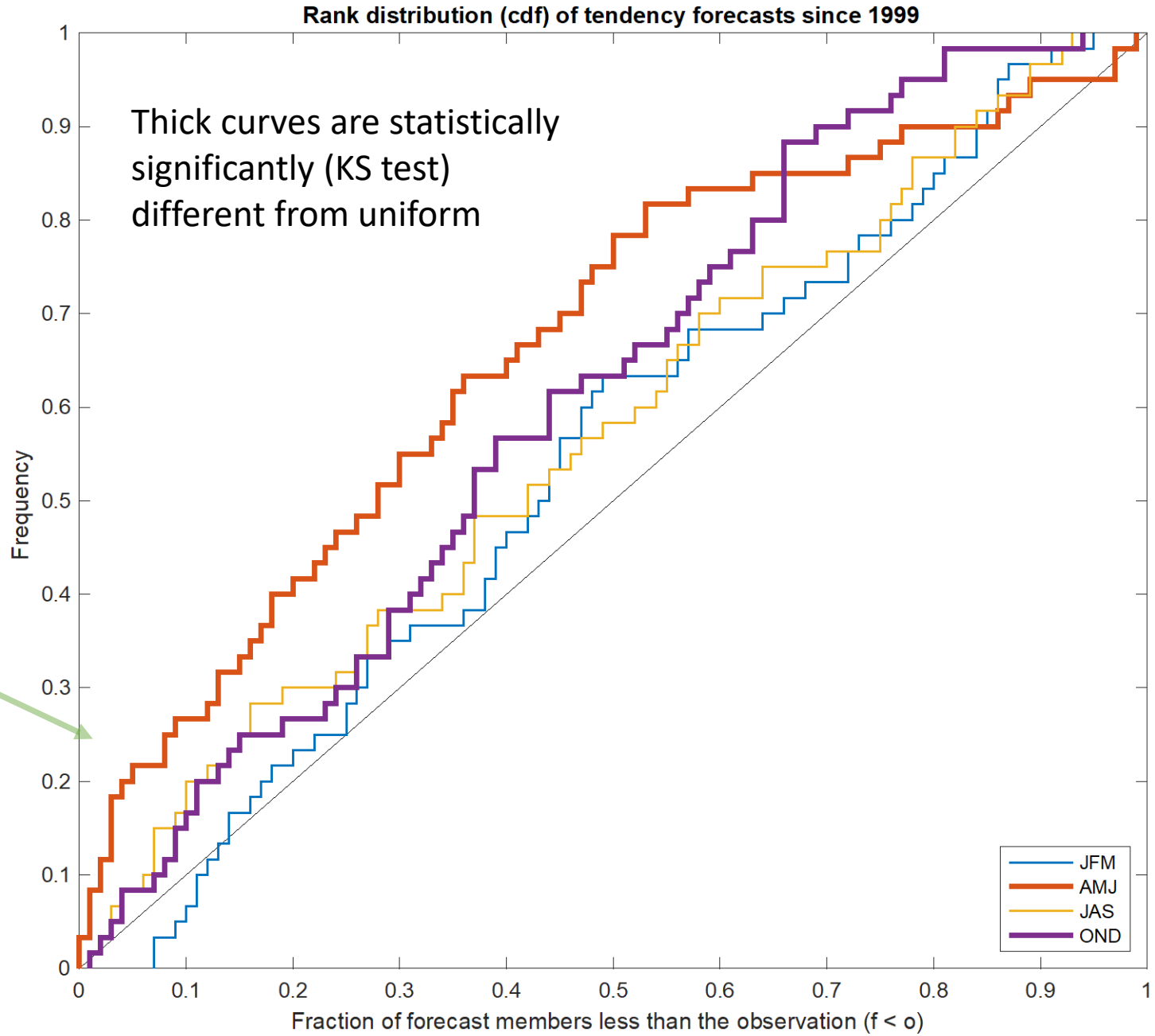




When the ensemble lies mostly above the observation, the fraction of ensemble members less than the observation is small.

The fraction of ensemble members less than the observation is small too often, which means the ensemble lies above the observation too often.





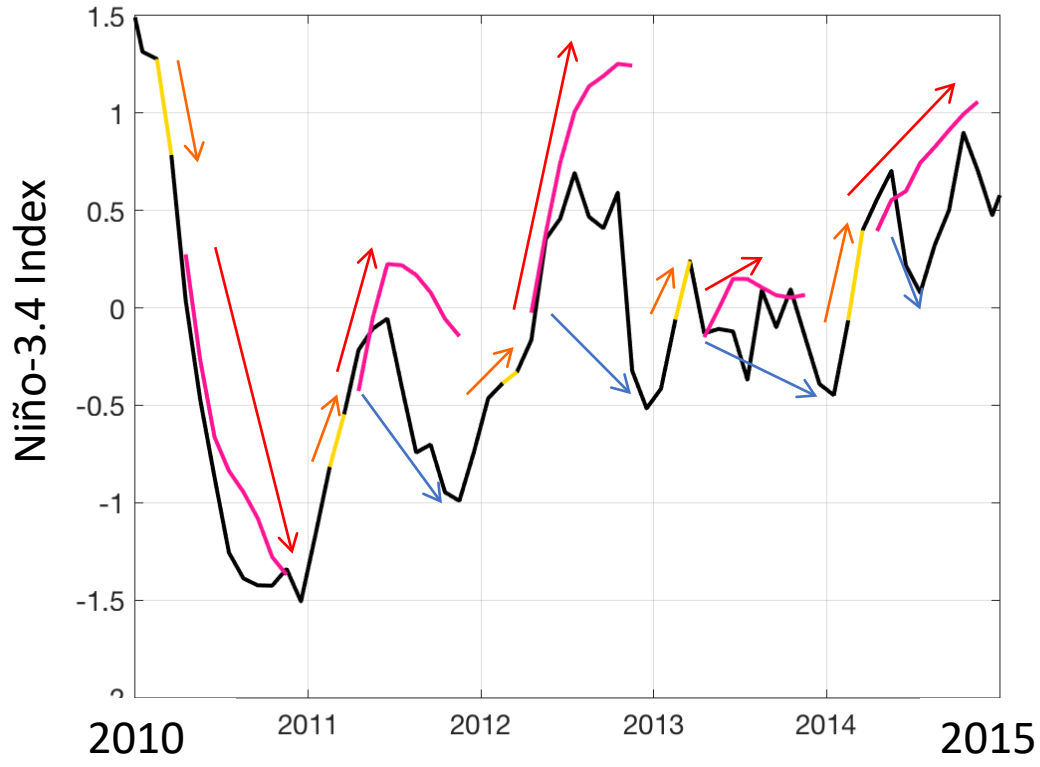
Key insight from ENSO forecasters Emily Becker and Michelle L'Heureux:

"Momentum"

In the last decade, we've experienced several "False Alarms" when the objective model guidance over-forecasted the chances of an ENSO event (especially forecasts made in **the spring**).

Niño-3.4 SST Index

NMME initialized in May (pink) and Observations (**black** & March-April in yellow)



Forecasters don't "feel" average skill. They feel individual forecasts, especially busts.

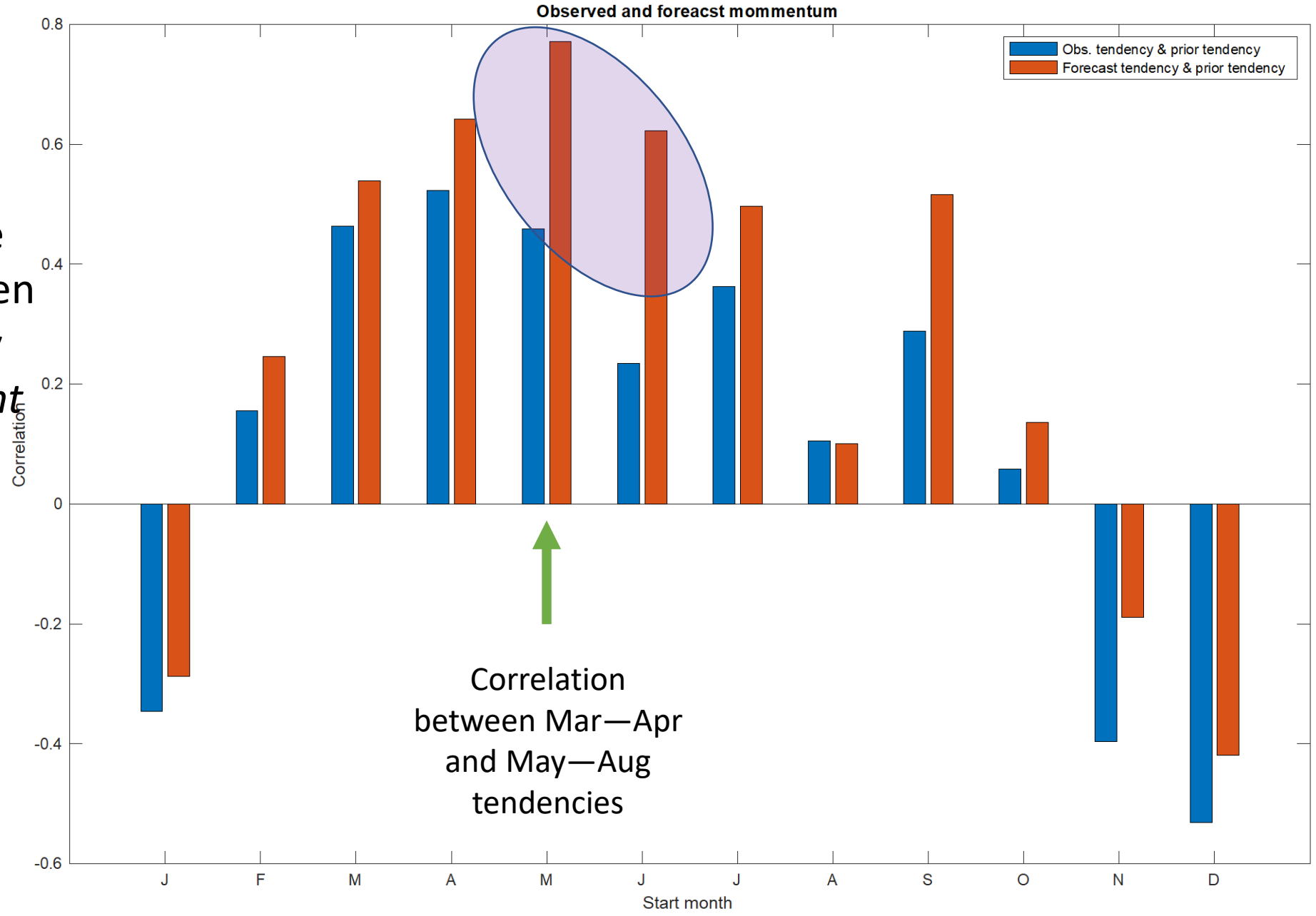
We want to objectively define these false alarms and have chosen the "Niño-3.4 tendency" to quantify them (not the only way).

Do models have too much “momentum” coming out of spring?

The forecast tendency follows the prior tendency too strongly

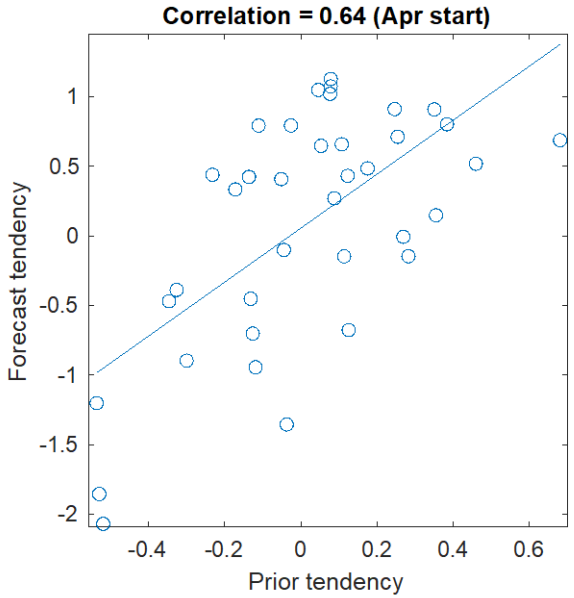
The past decade hasn’t been favorable to momentum-investing approaches. **ILLUSTRATION:** JASON SCHNEIDER FOR THE WALL STREET JOURNAL

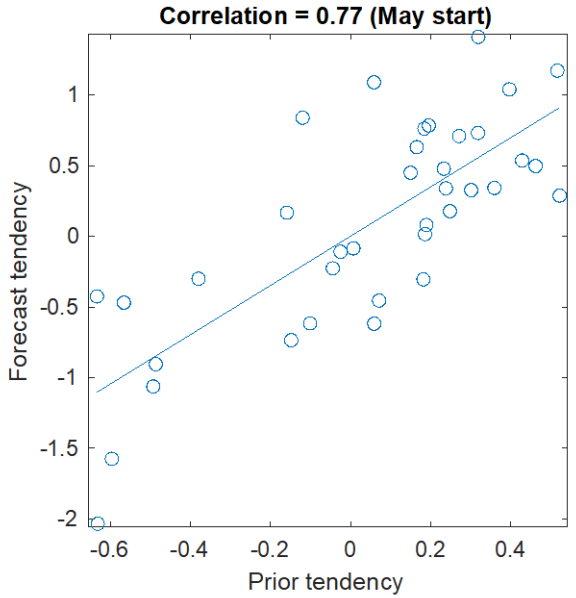
For each month of the year, we can define the **ENSO momentum** as the correlation between the *prior* tendency and the *subsequent* tendency

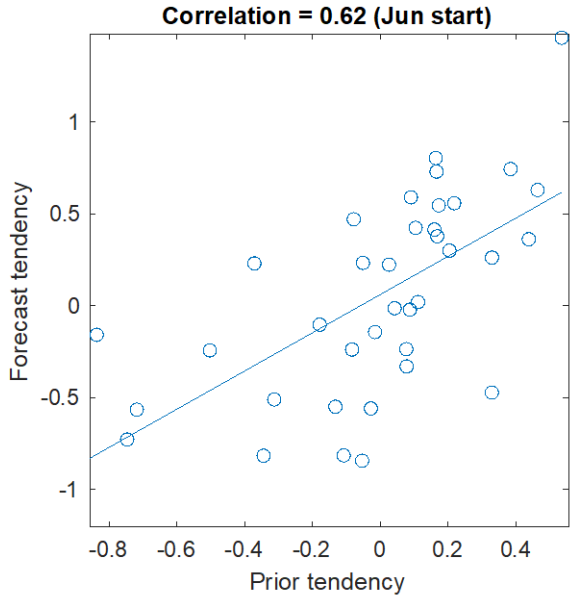


Observations
Forecasts

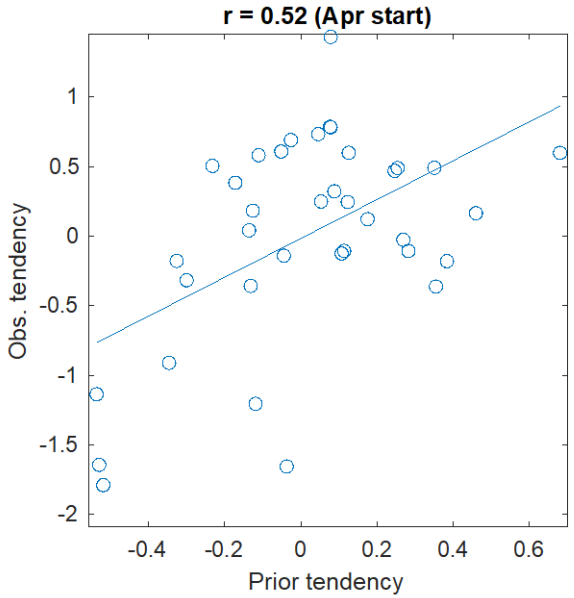
ENSO Momentum

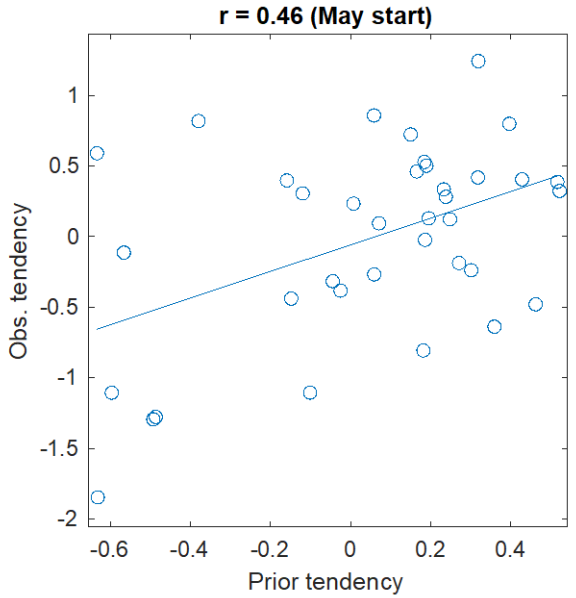


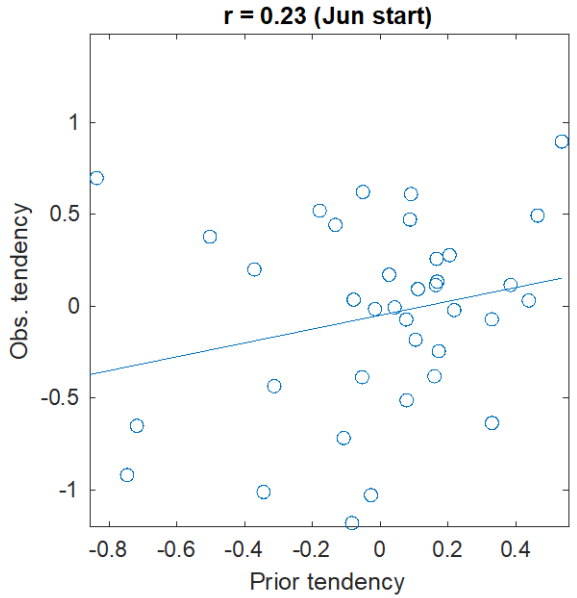




Ensemble-mean forecasts

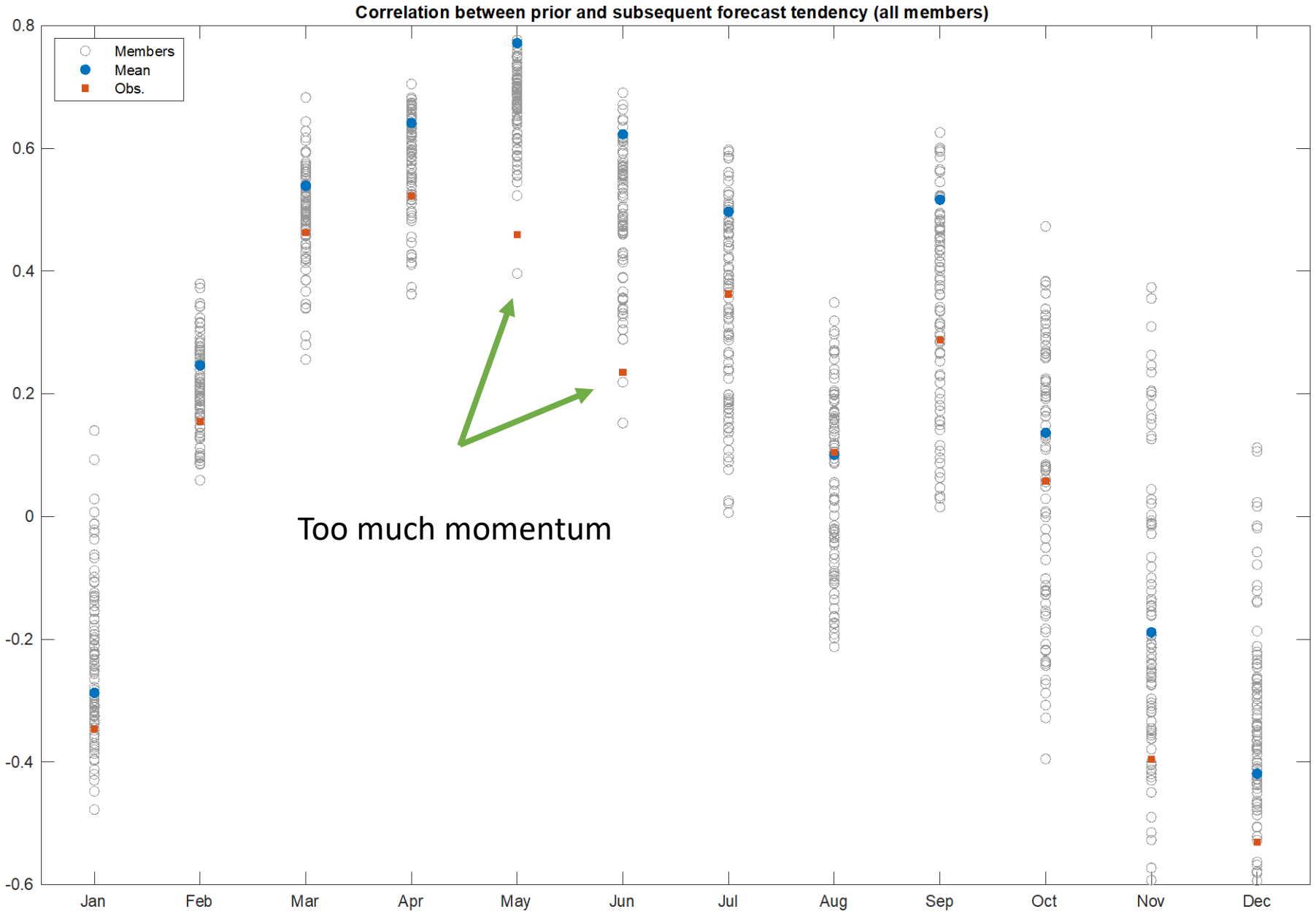






Observations

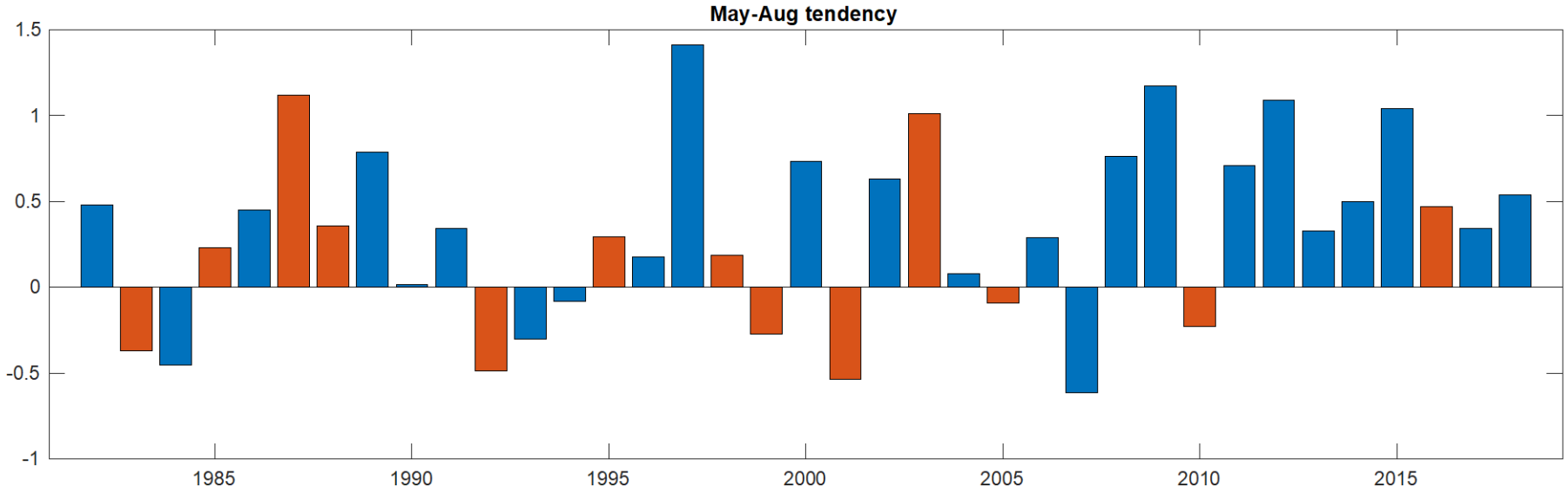
ENSO momentum in observations and ensemble members

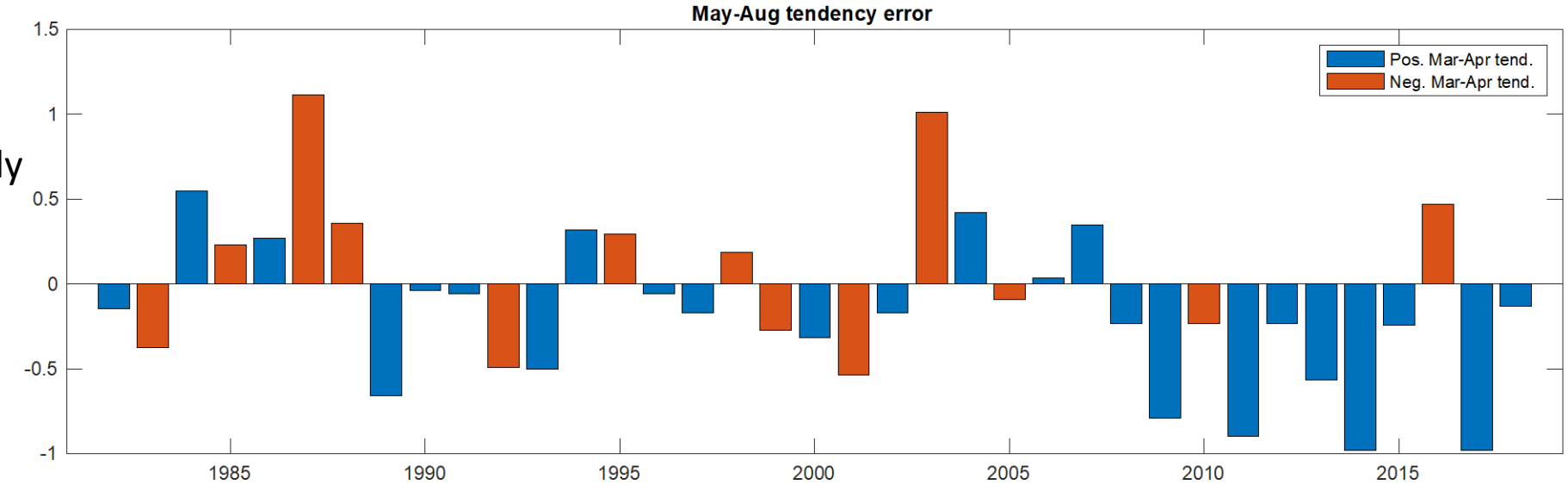


State-dependent forecast errors

1. Prior tendency is positive.
2. The forecast tendency is likely positive.
3. The observation not as likely to be positive.
4. The tendency error (obs-forecast) is likely to be negative

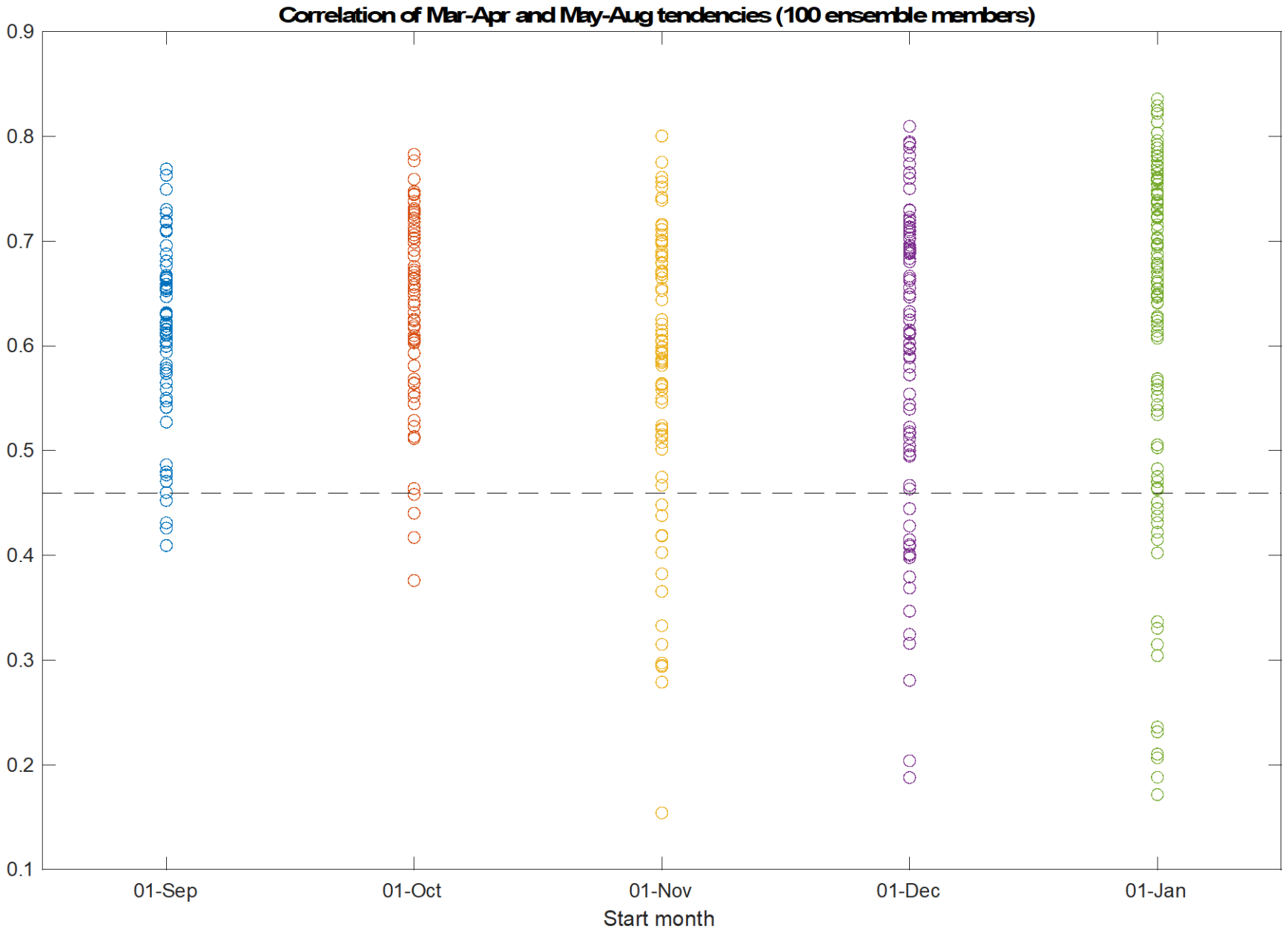
Correlation of prior and forecast error = -0.55





Initialization problems?

Earlier starts also have too much momentum. Initialization not entirely to blame.

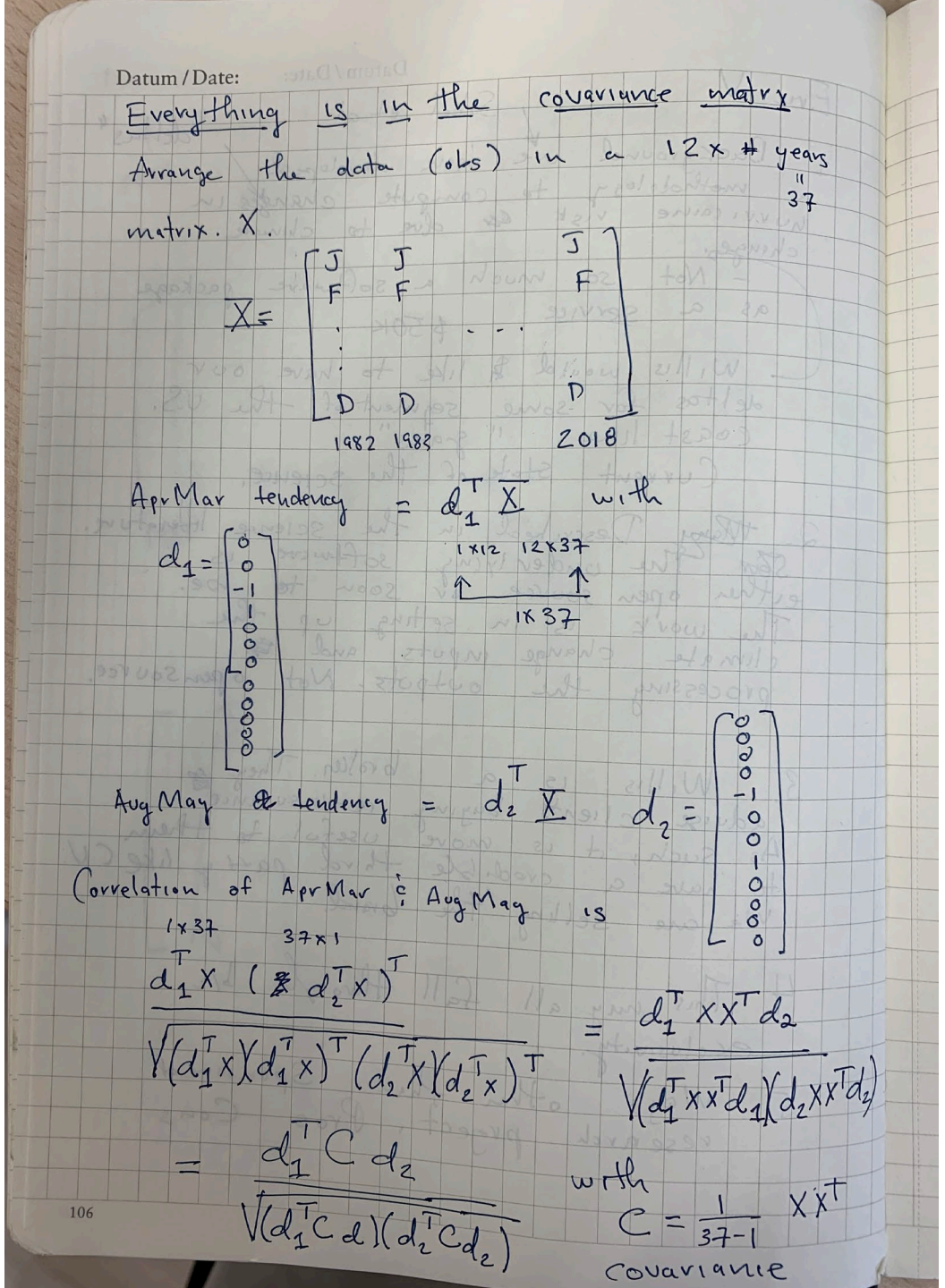


Could check in free running model.

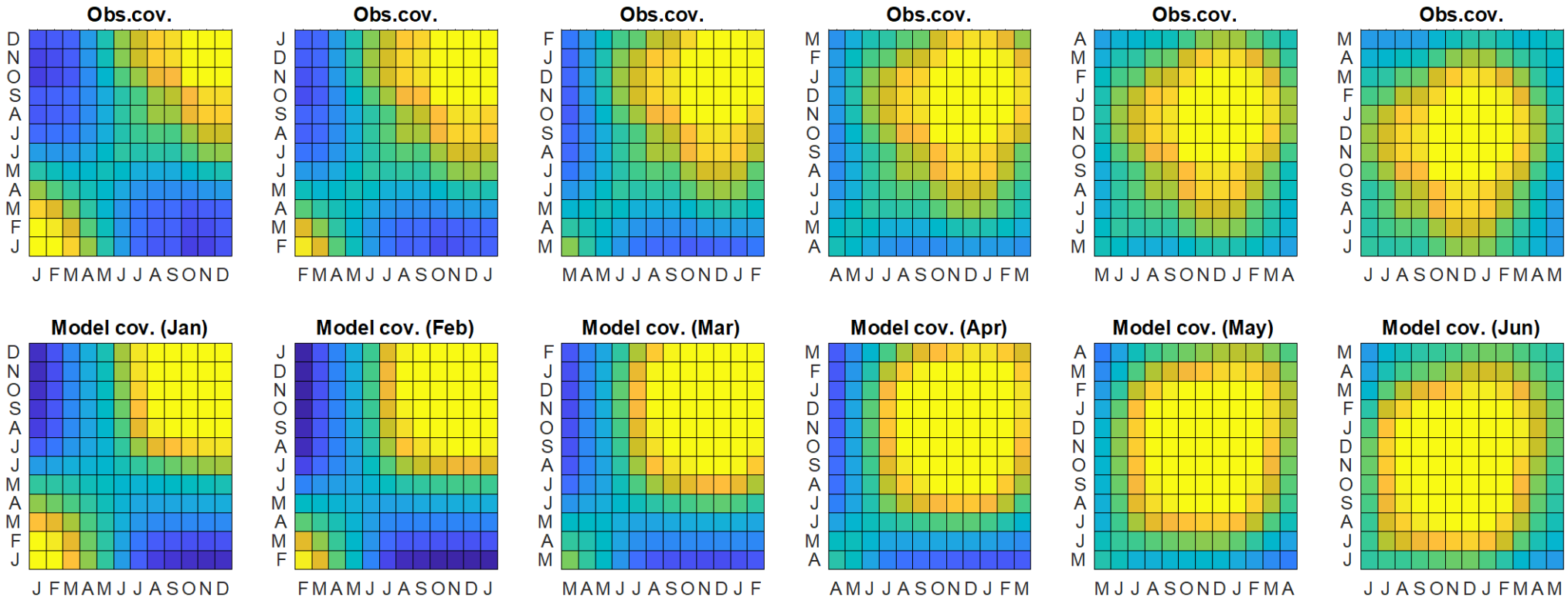
Linear algebra!

All of the time structure, including ”momentum” is in a covariance matrix

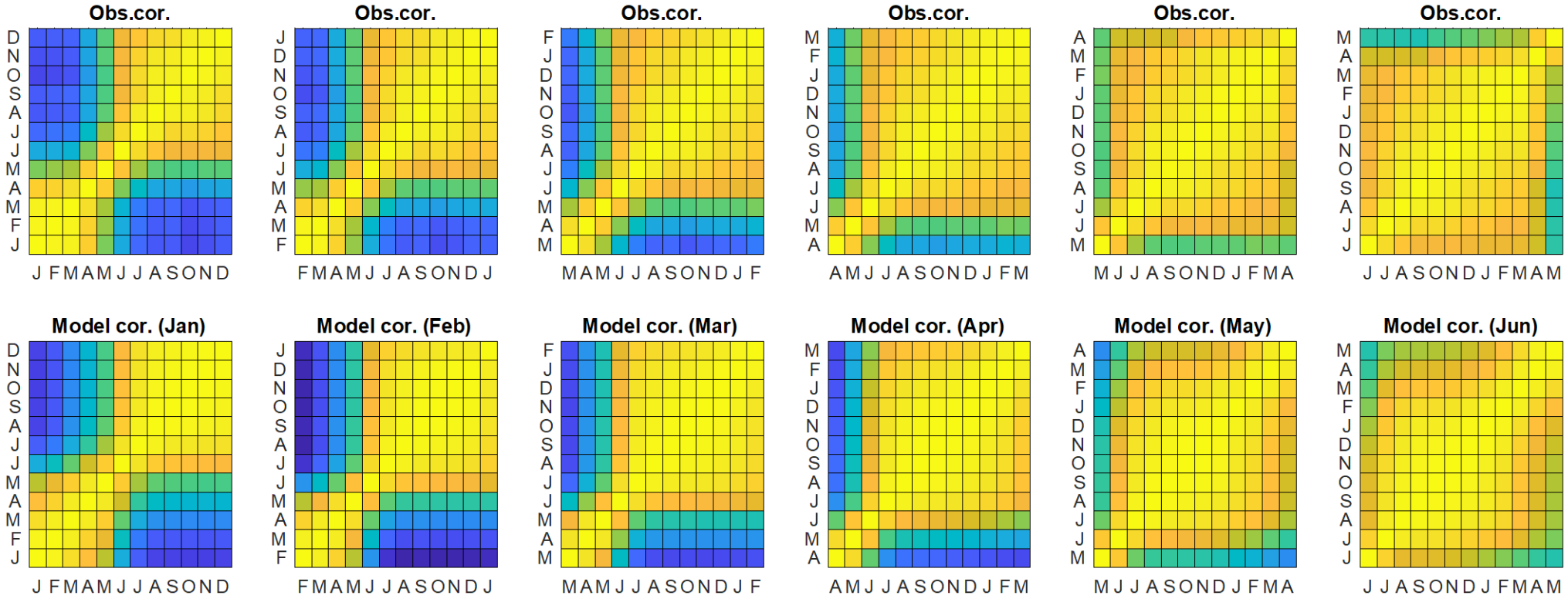
- Cut the Nino 3.4 times series into 12-month segments
- Put each segment into the column of a matrix X
  - X has 12 rows
    - Row 1 is all the January values
    - Row 2 is all the February value
    - And so on
  - The number of columns = # years
- $C = XX^T$ /(number of years -1)



We can look at covariance matrices …

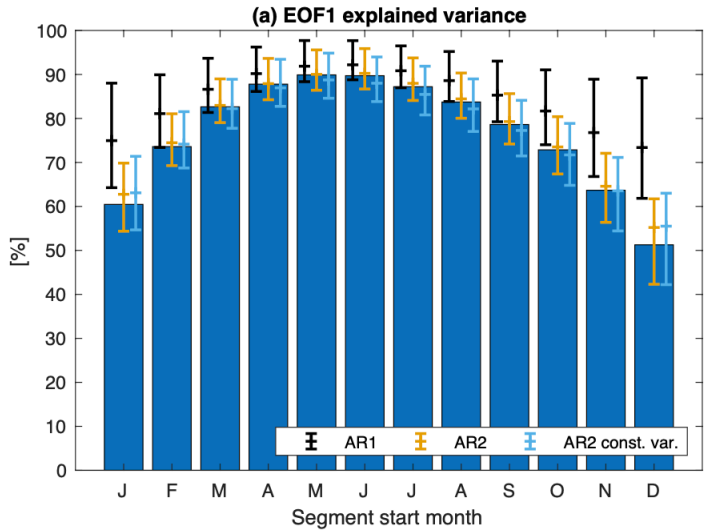


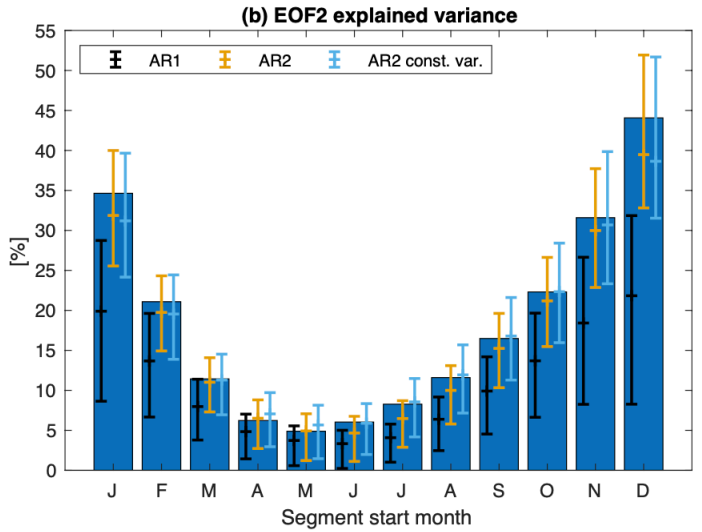
We can look at correlation matrices …

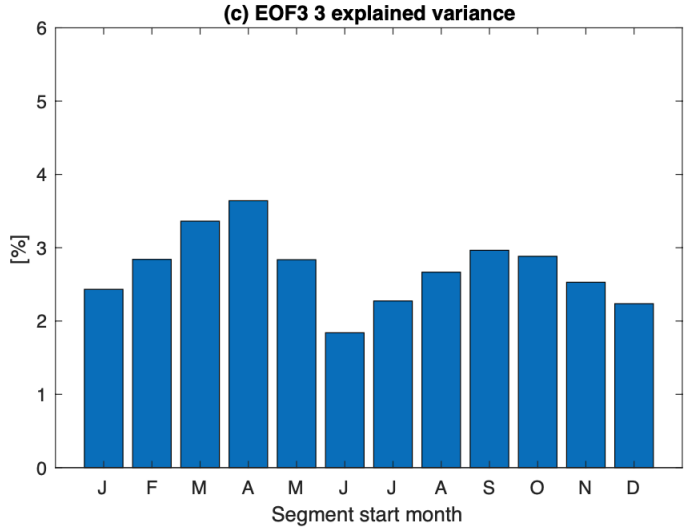


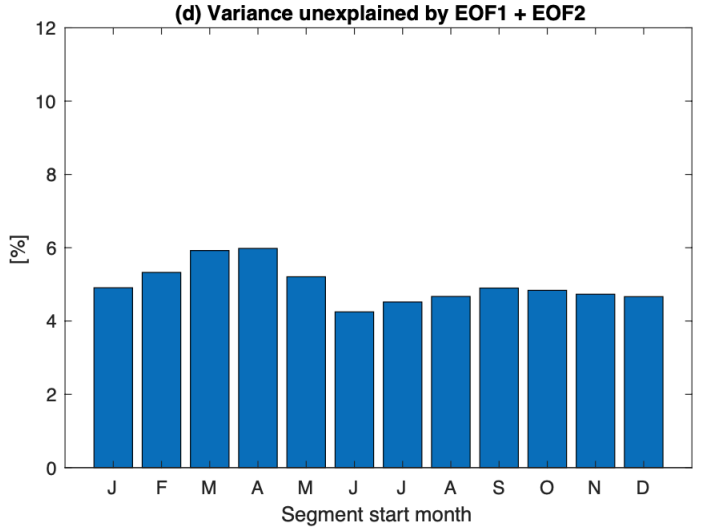
Luckily they are low-dimensional!

2 EOFs explain close to 95% of the variance

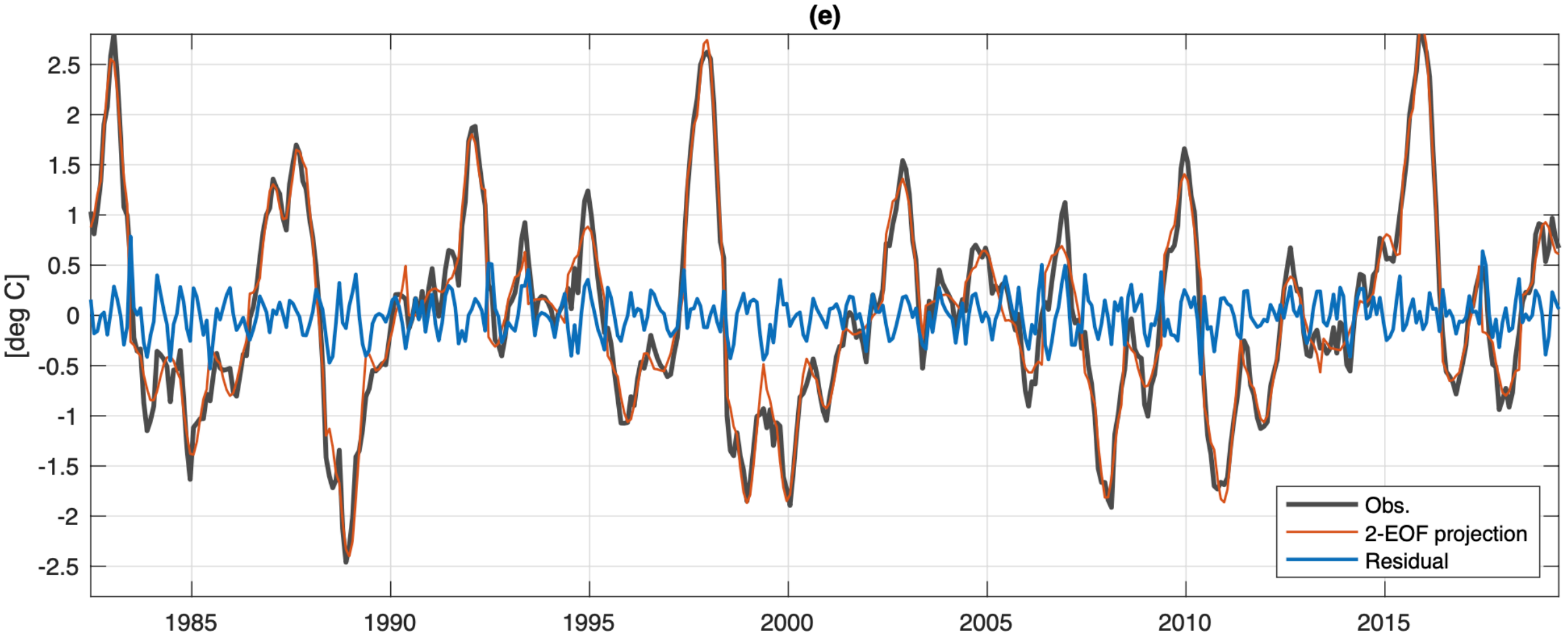




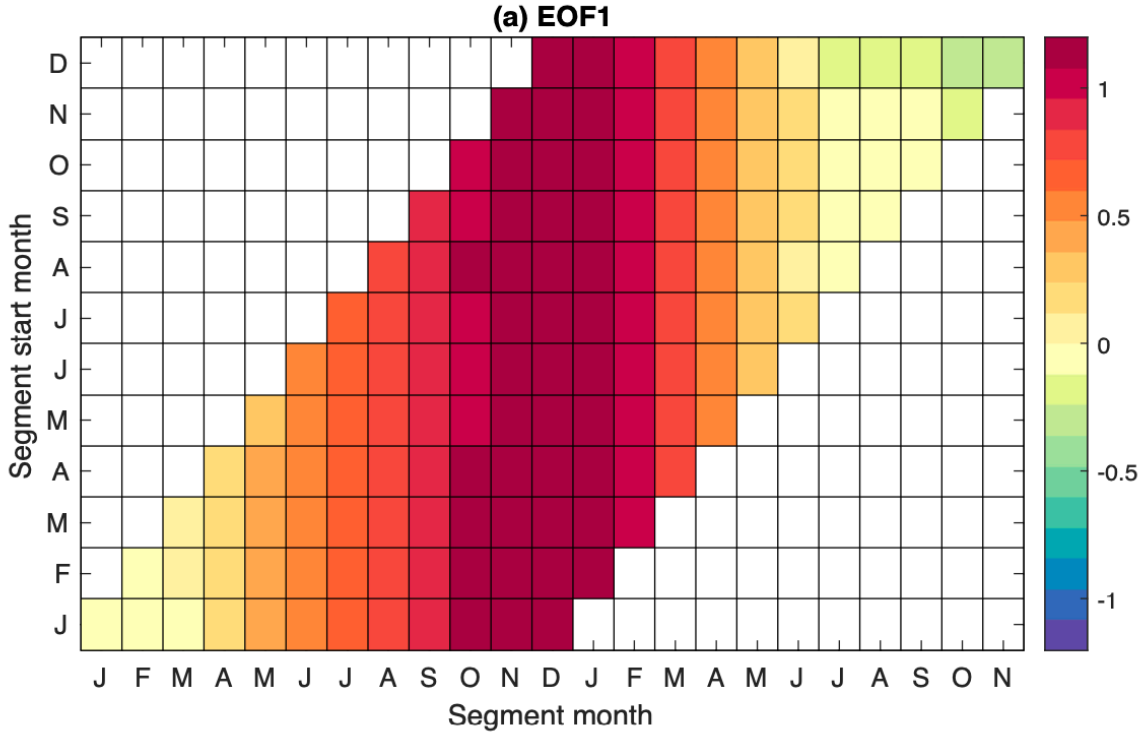


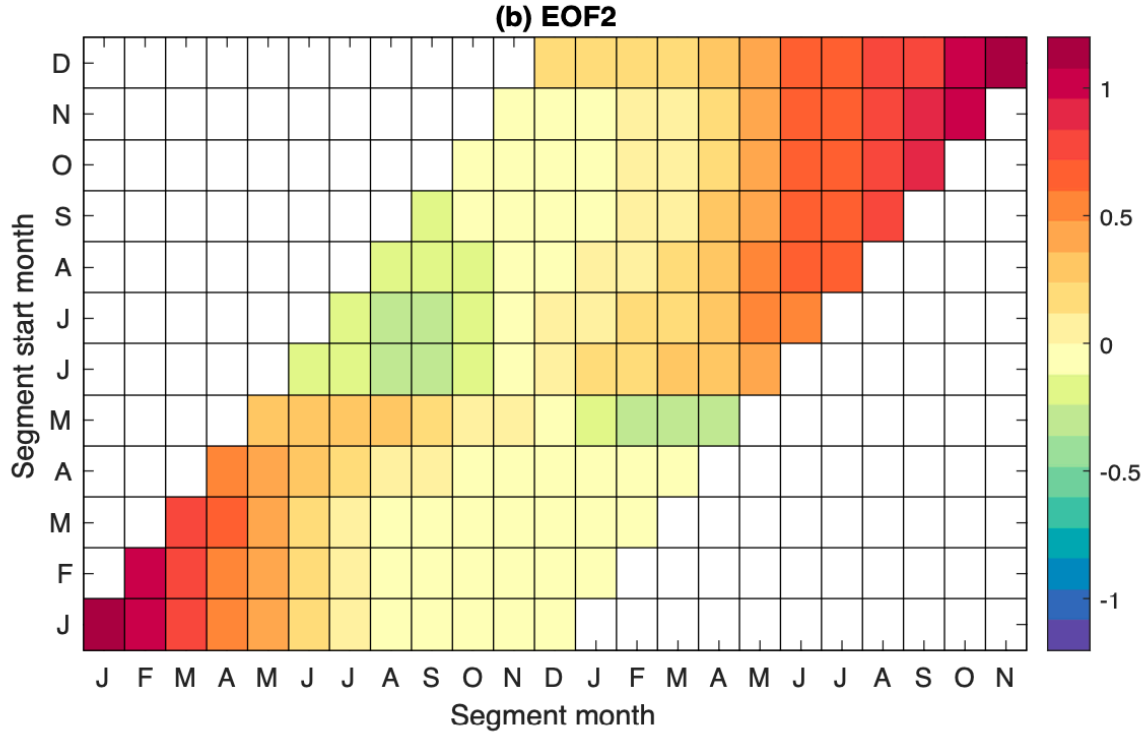


Approximation of Nino 3.4 using 2 EOFs

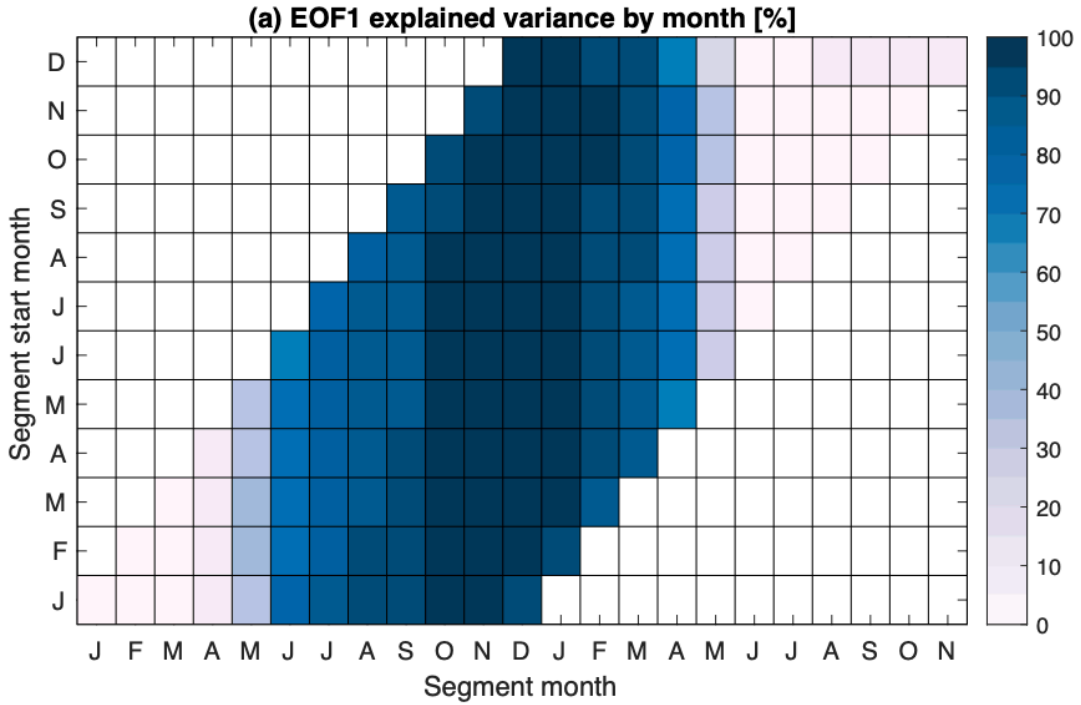


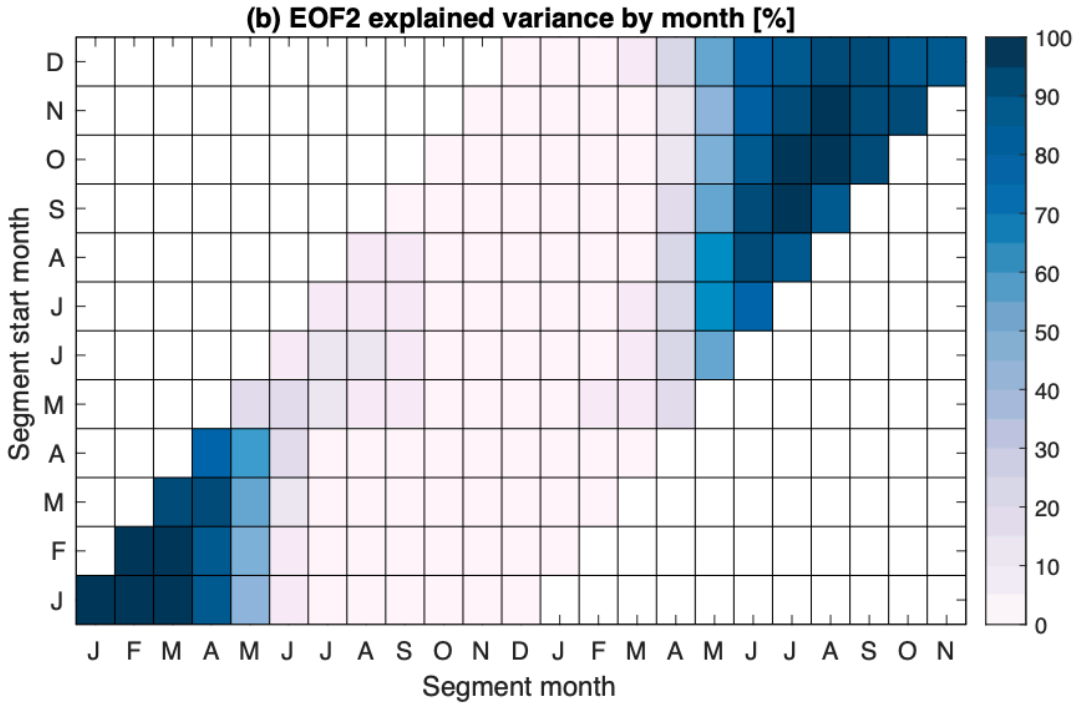
What do the EOFs look like?



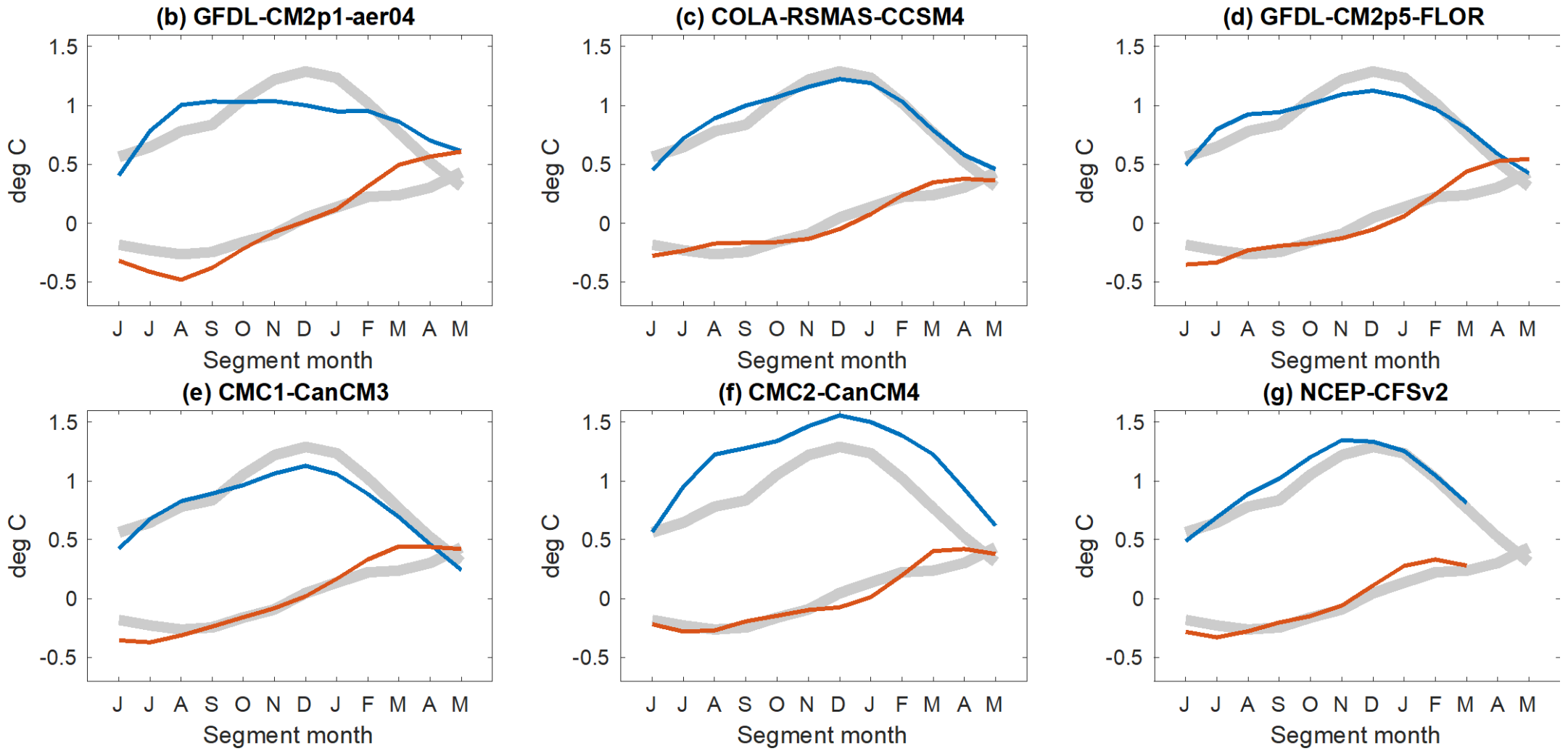


Variance explained

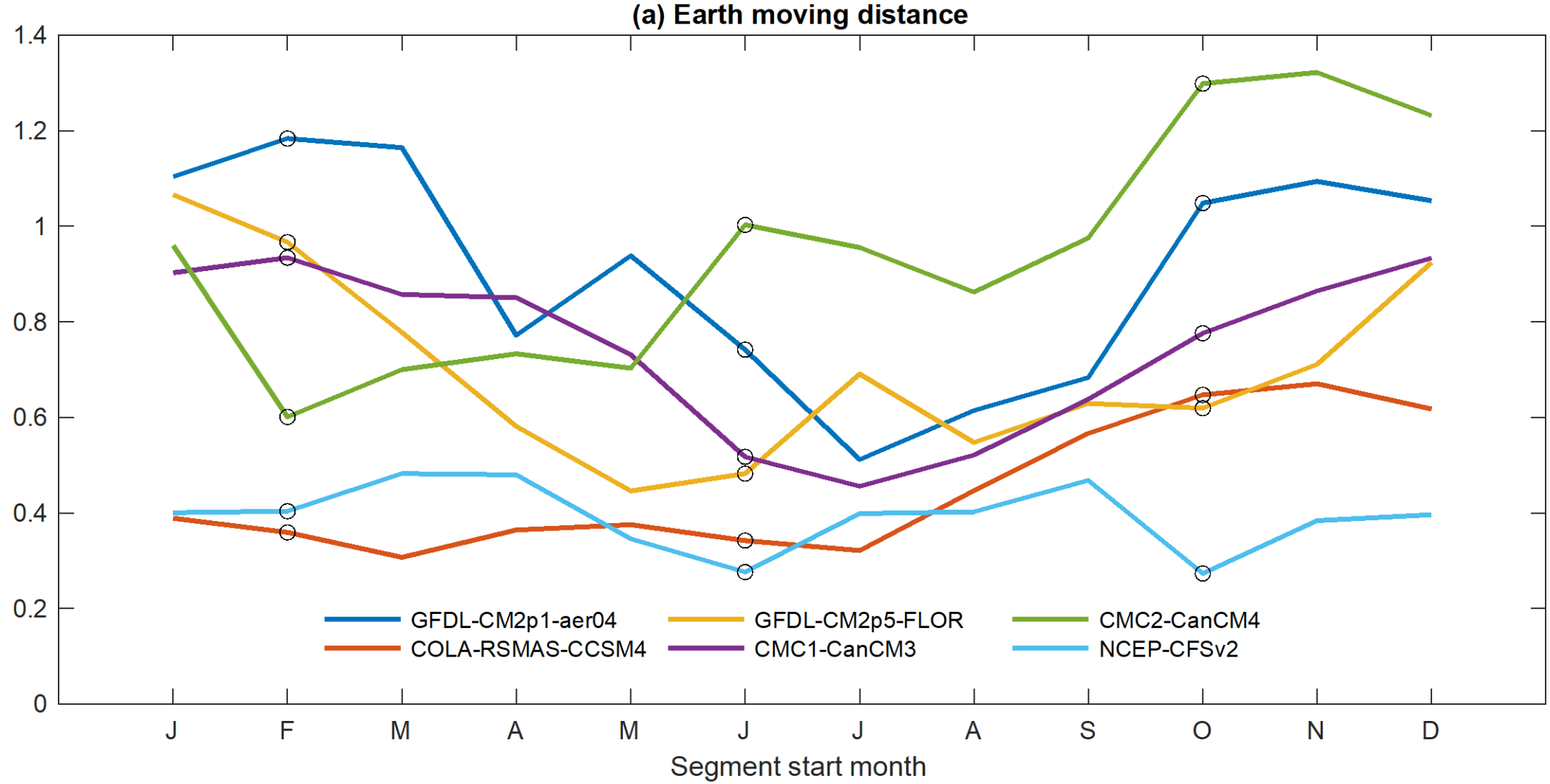




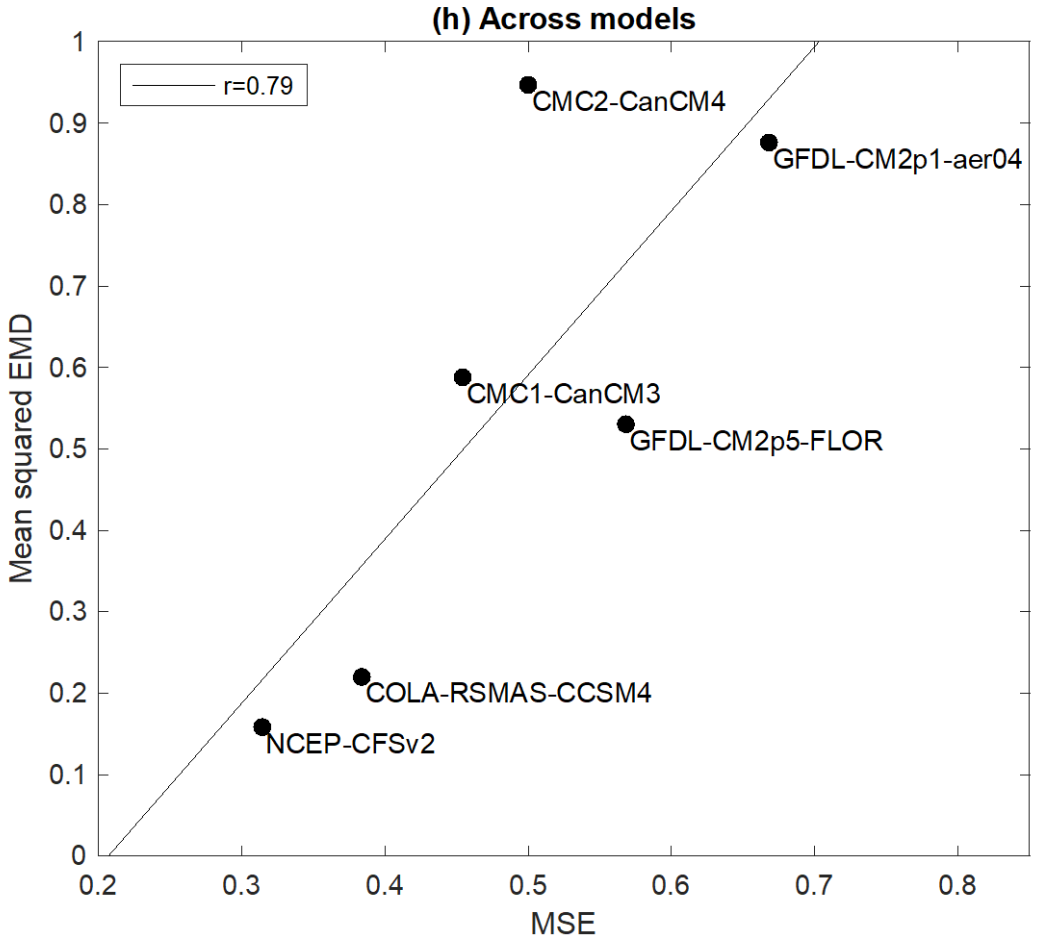
Back to forecasts. Compare EOF of forecasts and observations

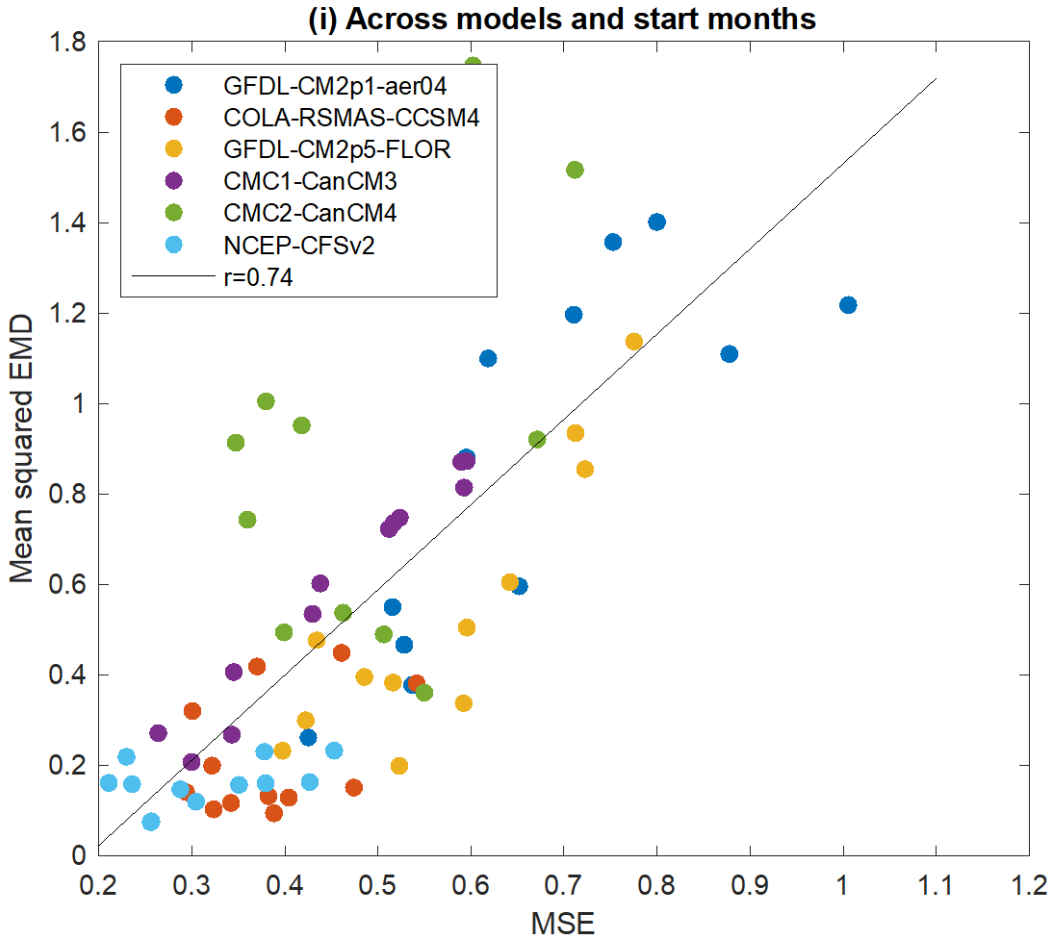


EMD measures distance between models and observed covariance

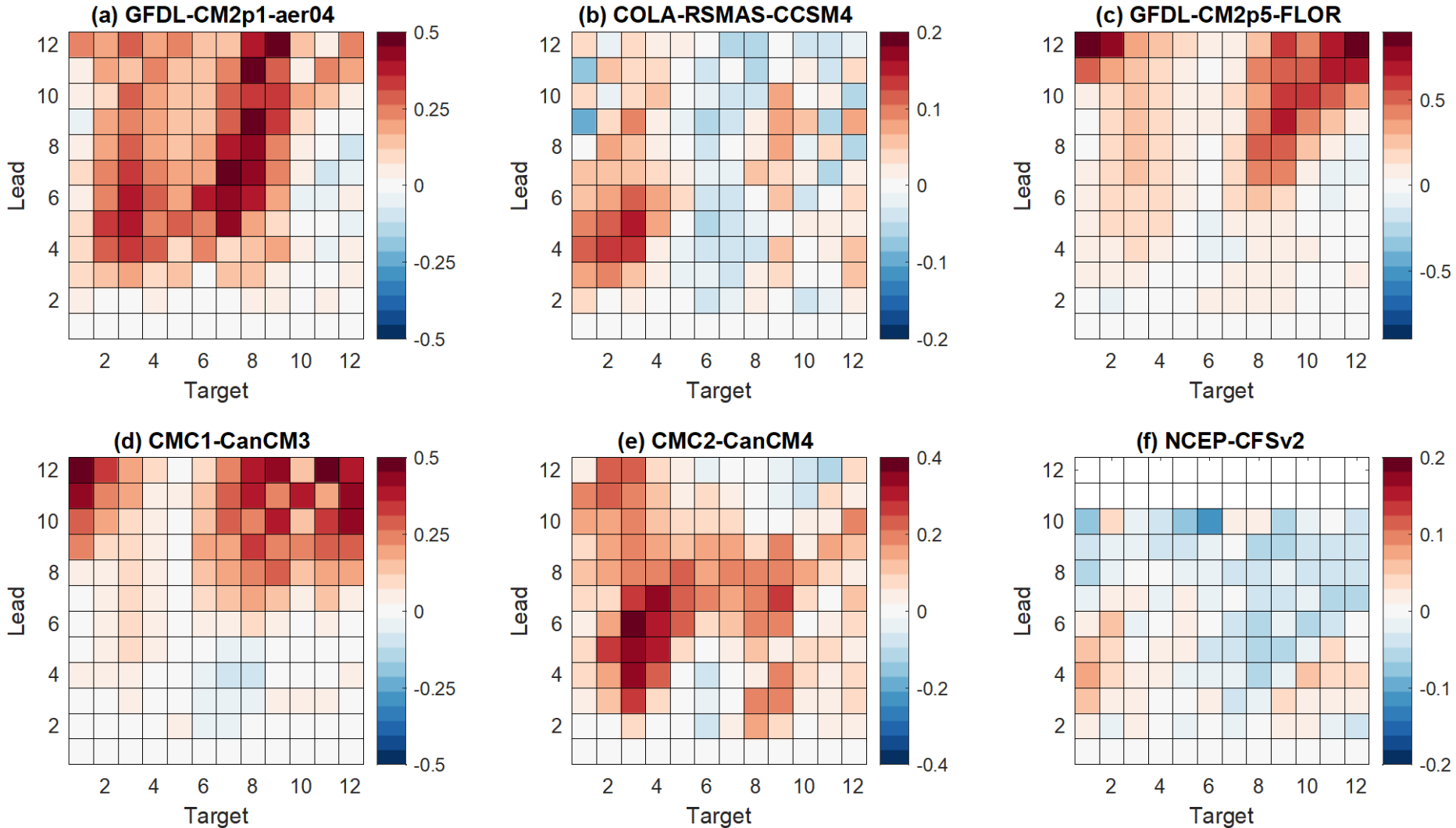


Models with better covariances have lower MSE





EOFs can be used in a regression to reduce mean-squared error



Summary

- Late-spring ENSO *forecasts errors* depend on the observed ENSO tendency before the forecast.
  - “Momentum” = ENSO tends to keep going in the same direction
  - Models have state-dependent errors
- Consequence: Late-spring ENSO forecasts are unreliable.
- Are model dynamics or initialization to blame?
  - Some evidence for dynamics being the culprit
- EOF analysis shows ENSO time structure is 2-dimensional
  - Models with better time structure have smaller errors.
  - EOF-based regression reduces forecast MSE