Assessing the Fidelity of Predictability Estimates

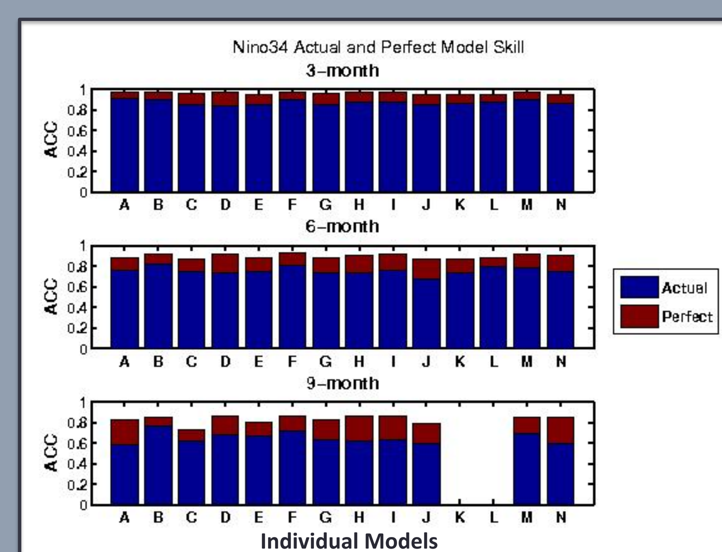
Kathy Pegion[1,2], Tim DelSole[1,2], Emily Becker[3], Teresa Cicerone[1]

[1]*Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University*
[2]*Center for Ocean-Land-Atmosphere Studies*
[3]*NOAA/Climate Prediction Center*

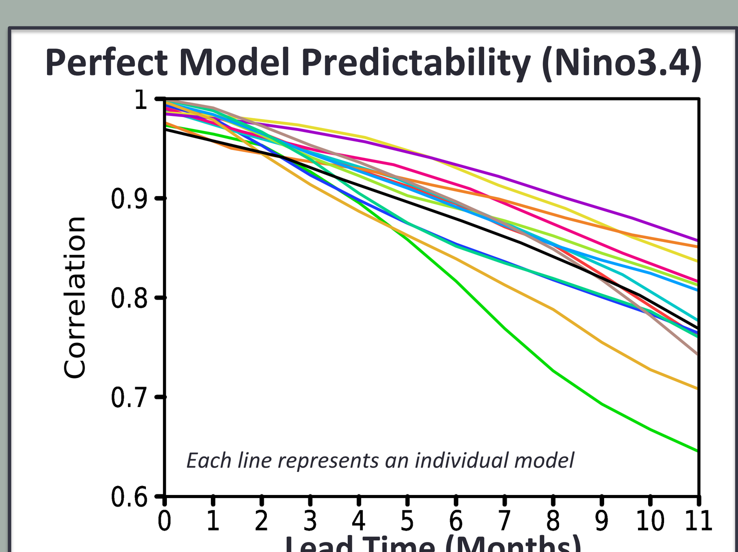
INTRODUCTION

Predictability is an intrinsic limit of the climate system due to uncertainty in initial conditions and the chaotic nature of the atmosphere. No matter how good models become, there will always be a limit to our ability to make predictions.

Estimates of predictability together with calculations of current prediction skill are often used to define the gaps in our prediction capabilities, inform future model developments, and set expectations for stakeholders.



PREDICTABILITY ESTIMATES

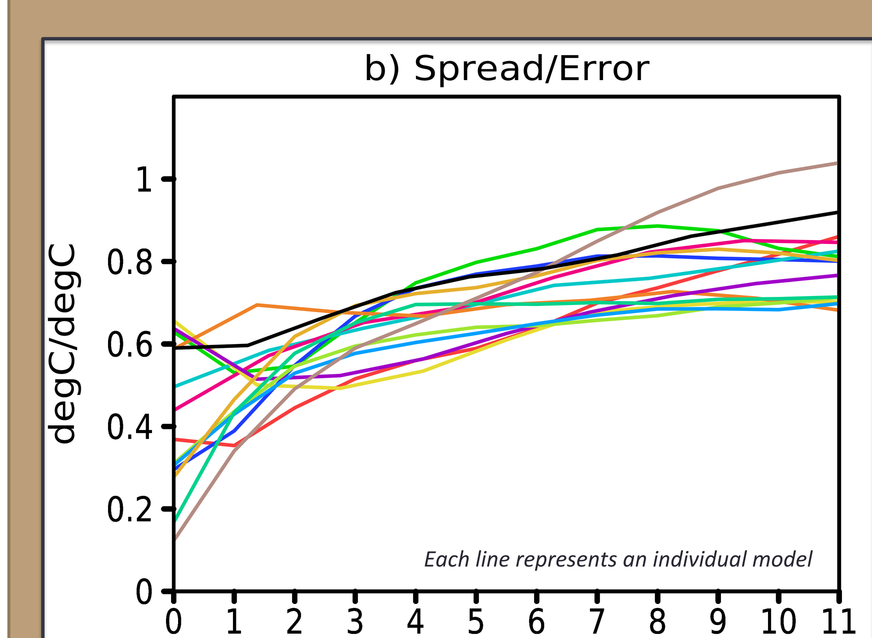


Perfect model predictability

- Used to assess how skillful a prediction system would be due to the growth of errors associated with uncertainties in the initial conditions.
- Calculated by withholding each member of an ensemble forecasting system in turn as the ``truth'' and calculating how well the ensemble mean of all other members forecasts the withheld member.

Different models give different estimates of predictability
How do we know which predictability estimate is most representative of the true predictability of the climate system?

METRICS

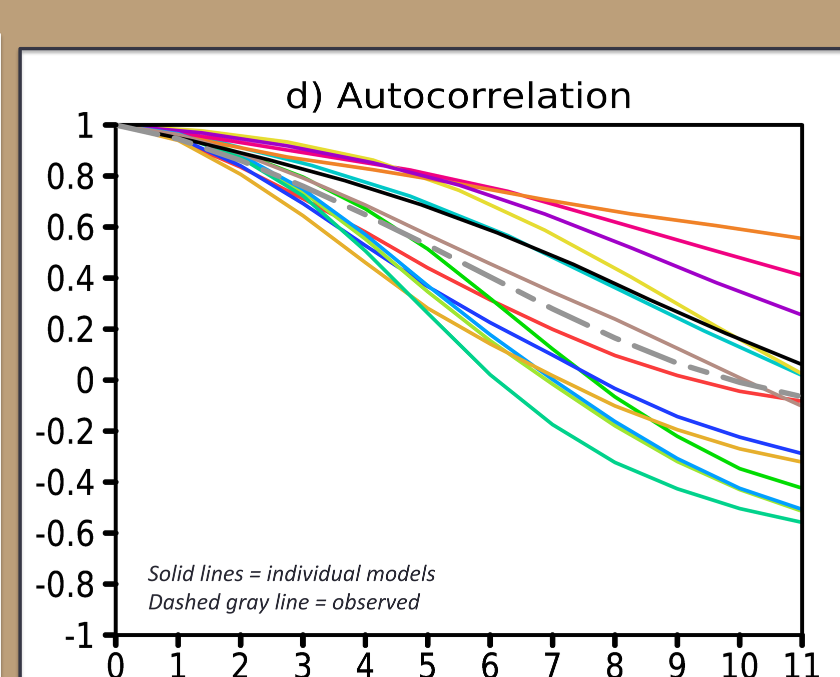


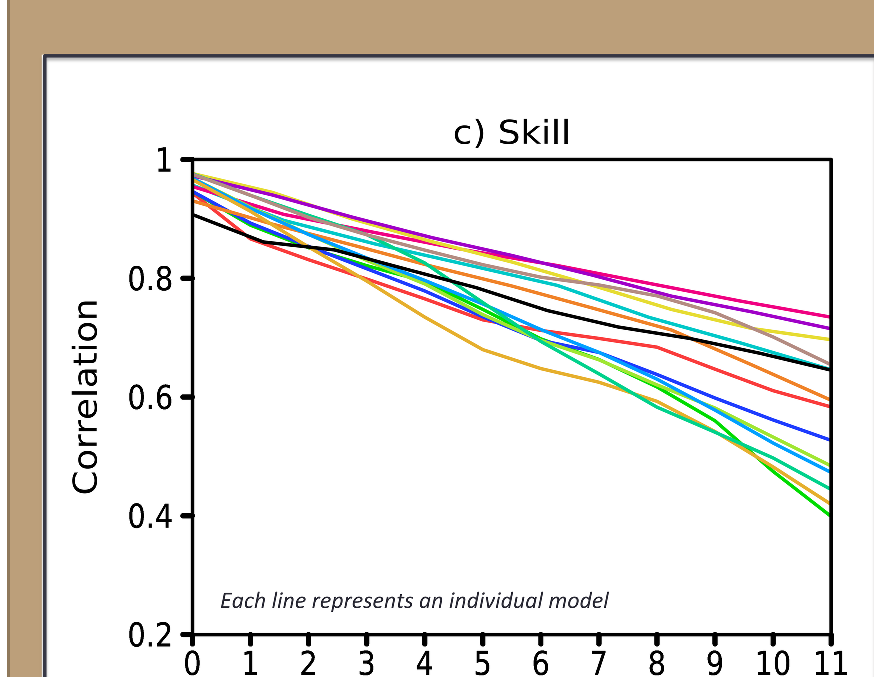
Spread/Error Ratio

- In a perfect ensemble prediction system, the spread-error ratio should be one.
- If the spread is underestimated relative to the error, then the ensemble is under-dispersive (over-dispersive), leading to over- (under-) estimates of predictability.
- All models have a spread-error ratio less than one except at the longest lead times. Given the uncertainties due to initial conditions and model errors, the spread is generally too small in these models to represent the forecast uncertainty (left).

Autocorrelation

- This metric provides a measure of persistence.
- If a model is more (less) persistent than the observations, then it seems intuitive that it would better (worse) predict itself, leading to predictability estimates higher (lower) than the true climate system.
- Figure to right shows the autocorrelation of the Nino3.4 index as a function of lead-time for each model's ensemble mean. The gray dashed line indicates the autocorrelation for the observations. Clearly, some models are more persistent than observations, while others are less persistent.





Skill

- Given that skill is a measure of how close an ensemble prediction system is to the truth on average over many cases, it seems reasonable to expect that the most skillful model is the model most similar to observations and therefore may have the most realistic predictability estimate.
- Figure left shows the actual skill of each model for the Nino3.4 index relative to observations. As with predictability estimates, there is also a wide range of skill, particularly as the lead time increases, ranging from about 0.65 to 0.85 at 6-months lead time.

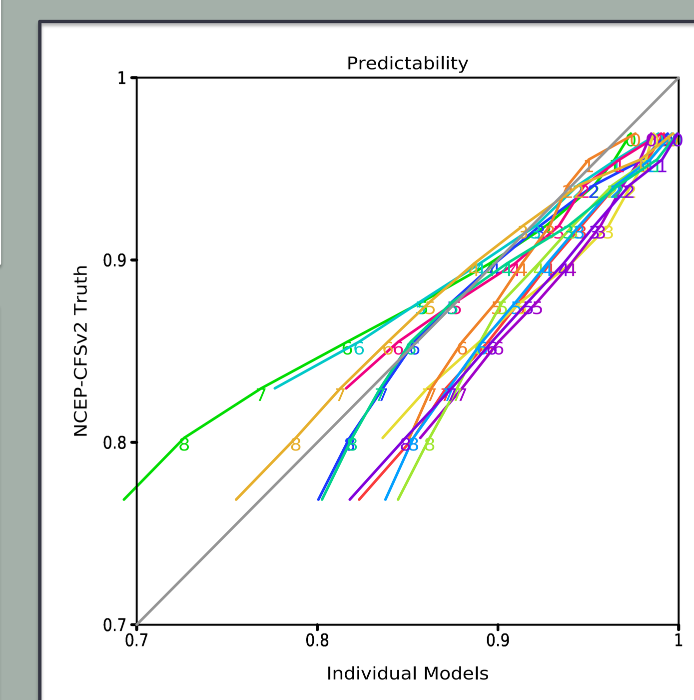
METHODOLOGY

Idealized framework using the North American Multi-model Ensemble (NMME) Re-forecast Database

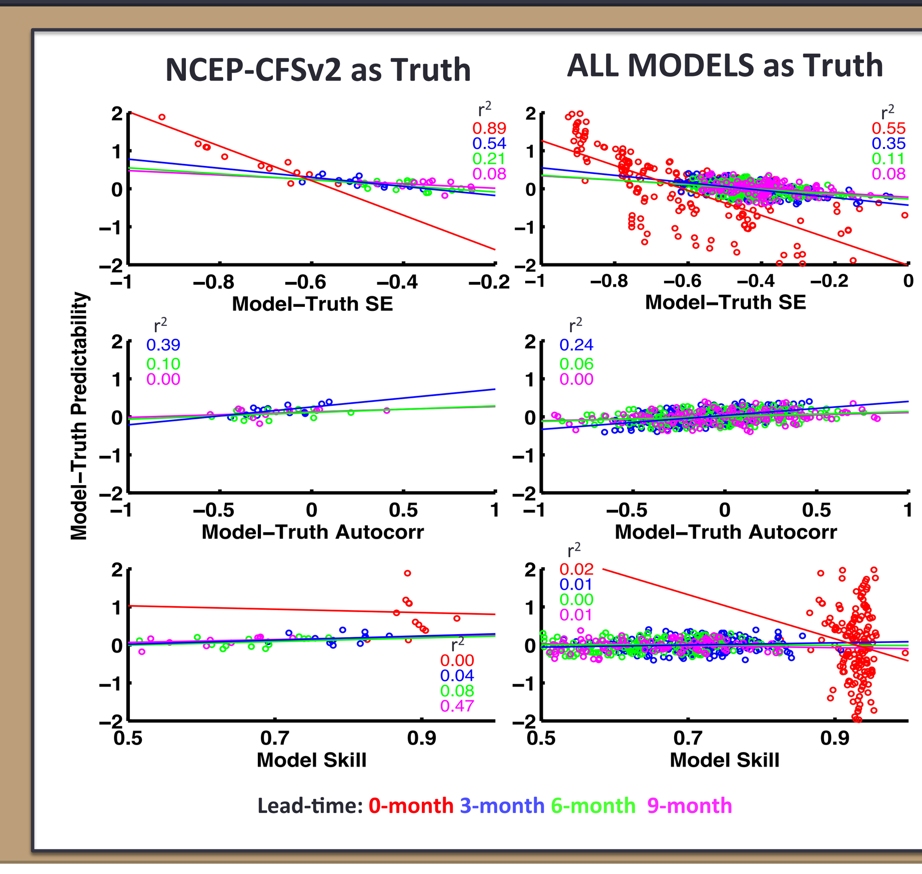
1. Take each model in turn as the truth and calculate "true" predictability
2. Compare how well all other models estimate this predictability
3. Apply metrics to see if they can provide any information about how well a model represents "true" predictability

Example

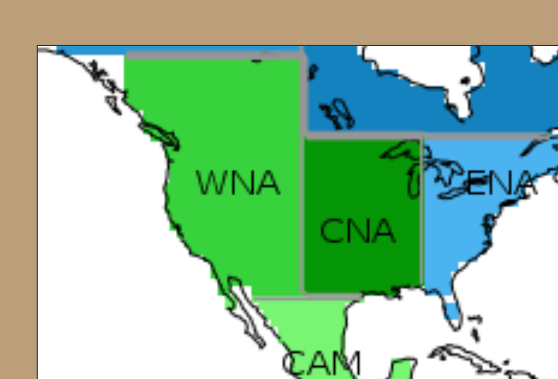
- Figure, right, shows the predictability estimates of each model relative to the NCEP-CFSv2 model selected as "truth".
- Most of the models have higher predictability estimates than the NCEP-CFSv2 model, although some have lower estimates than NCEP-CFSv2.
- Could we identify whether a model will have a higher, lower, or similar estimate of predictability to the NCEP-CFSv2 by looking at characteristics of the predictions relative to the NCEP-CFSv2? We apply the three metrics that represent characteristic of the prediction systems -- SE ratio, autocorrelation, and skill to test.

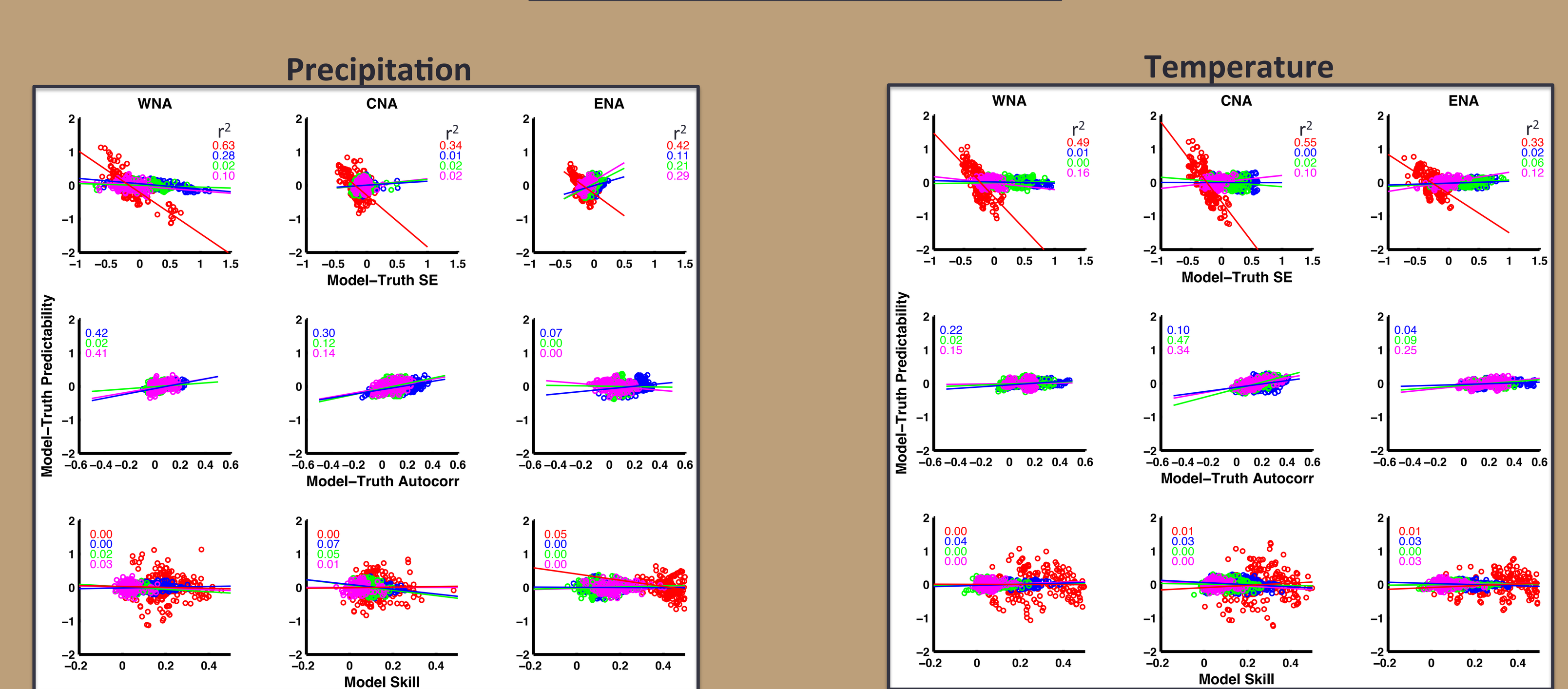


RESULTS: Nino3.4



RESULTS: US Temperature & Precipitation





CONCLUSIONS

☑Spread-Error Ratio

- But only at short lead-times
- Perhaps useful for weather and subseasonal timescales? (Future work)

☒Autocorrelation

Not robust across regions or lead-times

☒ Skill

Nothing useful here