Identifying and Assessing Gaps in Subseasonal to Seasonal Prediction Skill
Progress Report

1. General Information

Project Title: Identifying and Assessing Gaps in Subseasonal to Seasonal Prediction Skill
PI/co-PI names and institutions: Kathy Pegion, George Mason University
Report Year: Final
Grant #: NA15OAR4310072

2. Main goals of the project, as outlined in the funded proposal

- Systematically quantify estimates of the upper limits of predictability
- Assess similarities and differences between predictability estimates and understand the reasons for differences between them
- Compare predictability estimates with current skill to identify gaps in our prediction capabilities.

3. Results and accomplishments

Predictability is an intrinsic limit of the climate system due to uncertainty in initial conditions and the chaotic nature of the atmosphere. Estimates of predictability together with calculations of current prediction skill are used to define the gaps in our prediction capabilities, inform future model developments, and indicate to stakeholders the potential for making forecasts that can inform their decisions.

The true predictability of the climate system is not known and must be estimated, typically using a perfect model estimate from an ensemble prediction system. However, different prediction systems can give different estimates of predictability (Figures 1 and 2). Which estimate of predictability is most representative of the true predictability of the climate system?

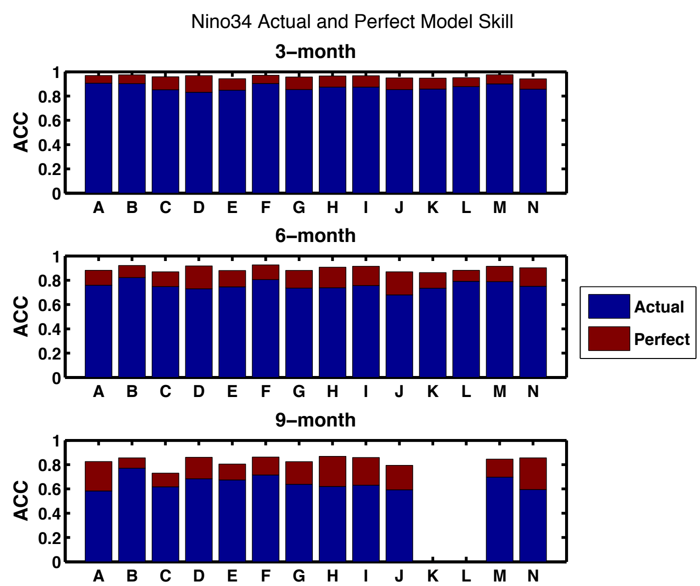

Figure 1: Skill of Nino3.4 index predictions (blue) and perfect model predictability estimates (red) for 3-month (top), 6-month (middle), and 9-month (bottom) lead times. Letters indicate individual models from the NMME.

To answer this question, it is necessary to identify some metrics associated with the characteristics of a prediction system relative to the characteristics of the observed system that can provide some insight – the spread-error ratio (SE), the autocorrelation, and the prediction skill. Examples of each of these metrics for the Nino3.4 from the fourteen models in the NMME database are shown in Figure 2.

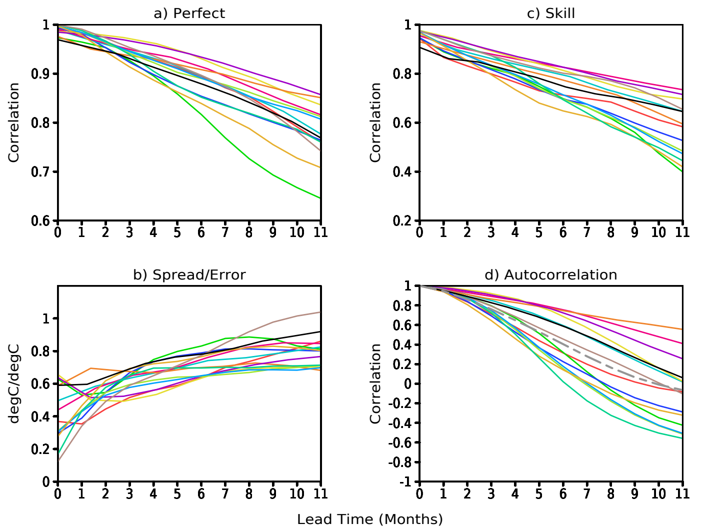

Figure 2: Perfect model predictability estimates (a), SE ratio (b), anomaly correlation prediction skill (c), and autocorrelation (d) for the Nino3.4 index for each model. Different colored lines indicate the different models and are consistent between panels.

Using monthly data from the North American Multi-model Ensemble re-forecast database, we quantify whether these metrics accurately indicate a model's ability to estimate predictability. Each of the metrics is shown in Figure 3 as a scatterplot relative to errors in estimating predictability of the ``truth'' (y-axis). This figure demonstrates that for Nino3.4 the models typically have too small of a spread relative to their error. It also shows that for 0- and 3-month lead times, as the SE ratio gets closer to the true SE, errors in estimating predictability are reduced. This is quantified by linear regression lines and corresponding r^2 values of 0.89 and 0.54, respectively. At longer lead-times (e.g. 6- and 9-months), there does not appear to be a relationship between errors in SE and errors in estimating predictability. Similar figures are shown for the autocorrelation (middle panels) and skill (bottom panels). If higher (lower) autocorrelation than the truth were related to overestimates (underestimates) of predictability, then we would expect to see a linear relationship in the scatterplot. For autocorrelation, no significant relationship appears to exist. If the skill is related to errors in predictability, we would expect higher (lower) skill to be related to lower (higher) errors in predictability. No such relationship is evident.

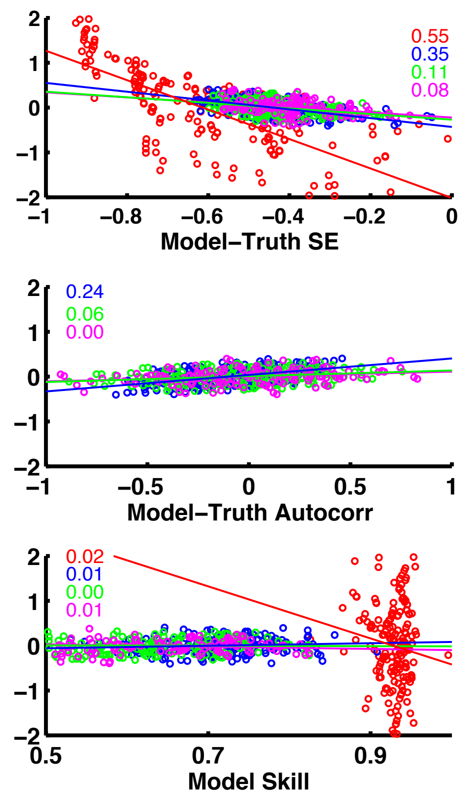

Figure 3: Scatterplots of the errors in the SE ratio (top), errors in autocorrelation (middle), and skill (bottom) for each individual model versus errors in estimating predictability of Nino3.4 when all models are used as truth.

We also test these metrics for temperature (not shown) and precipitation (Figure 4) over land. Figure 4 shows the corresponding scatterplots for each metric for the regions of western north America (WNA), central north America (CNA) eastern north America (ENA). The results are consistent with those for Nino3.4. Errors in the spread-error ratio are related to errors in estimating predictability at the shortest lead-times, while skill is not related to predictability errors. The relationship between errors in the autocorrelation and errors in estimating predictability varies by lead-time and region.

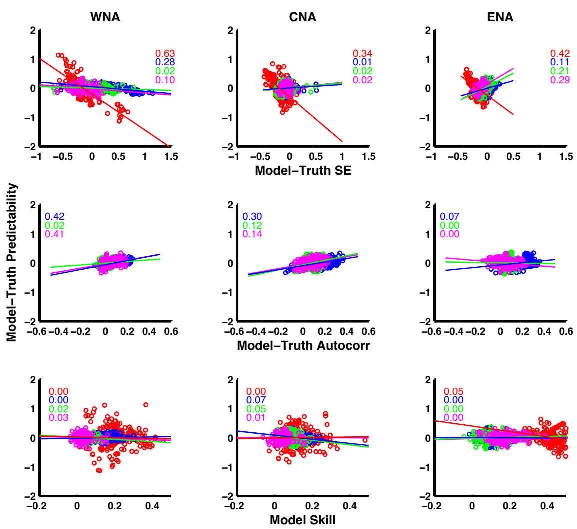

Figure 4: Scatterplots of the errors in the SE ratio (top), errors in autocorrelation (middle), and skill (bottom) for each individual model versus errors in estimating predictability of precipitation over land for all models as truth in the WNA (left), CNA (middle) and ENA (right) regions. Errors are calculated as model minus truth. For differences in correlations (i.e. y-axis in all panels and x-axis in middle and bottom panels) differences are between Fischer z-transformed values. Colors indicate lead times of 0-month (red), 3-month (blue), 6-month (green), and 9-month (cyan). Lines are the least squares fit through the corresponding points. Numbers indicate the r^2 values of the least-squares fit.

Our results demonstrate that none of these metrics provide a robust measure of the fidelity of predictability estimates in our idealized framework for Nino3.4. This means that none of them could be used to identify which models provided predictability estimate most similar to the model selected as truth. Given that these metrics are not robust in our idealized framework, we conclude that they cannot be used to select a model's predictability estimate as more or less realistic than another model's estimate of predictability in the real world.

The fact that the spread-error ratio appears to be related to errors in estimating predictability at the shortest lead-times indicates that this relationship may be useful for assessing predictability estimates at subseasonal timescales. This evaluation is ongoing using daily output from the NMME Phase 2 data for MJO indices, temperature and precipitation.

This work has been presented at the CLIVAR Open Science Conference, Sep 2016, Qingdao, China, a MAPP Webinar, March 2017, the WGNE Workshop on systematic errors, Montreal, Canada, June 2017, and the NMME/SubX Science Meeting, College Park, MD, Sep 2017.

4. Highlights of Accomplishments

- Evaluated perfect model predictability estimates for Nino34 and global temperature, and precipitation in all NMME models
- Evaluated the spread-error ratio, autocorrelation, and forecast skill of Nino34 and global temperature and precipitation in all NMME models.
- Demonstrated that there is no quantitative relationship between errors in estimating predictability and skill or errors in spread-error ratio and autocorrelation. The exception being SE ratio at short lead times.
- Demonstrated that there is not a clear way to determine if one model's predictability estimate is more realistic than another model.
- Paper in press in Climate Dynamics special issue on NMME Evaluation

5. Transitions to Operations

None to report

6. Publications & Presentations from the Project

Pegion, K., DelSole, T., Becker, E. and Cicerone, T. Clim Dyn (2017). https://doi-org./10.1007/s00382-017-3903-7

CLIVAR Open Science Conference, Sep 2016, Qingdao, China
MAPP Webinar, March 2017
WGNE Workshop on systematic errors, Montreal, Canada, June 2017
NMME/SubX Science Meeting, College Park, MD, Sep 2017

7. PI Contact Information

Kathy Pegion
kpegion@gmu.edu