Collaborative Research: Separating Forced and Unforced Decadal Predictability in Models and Observations

Investigator:	Timothy DelSole
Recipient Name:	George Mason University
Award Number:	ER65095
Award Period:	1 September 2010 - 31 August 2015 (with no-cost extensions)
Reporting Period:	Award period
Program Office:	Biological and Environmental Research Program (BER)
Program Manager:	Renu Joseph
Participants:	Timothy DelSole, Michael Tippett Liwei Jia, Emerson LaJoie, Xiaoqin Yan

This report is a final report of the accomplishments of the research grant "Collaborative Research: Separating Forced and Unforced Decadal Predictability in Models and Observations" during the period award period 1 September 2010- 31 August 2015.

1 Summary of the Basic Research Plan

The purpose of the proposed research was to identify unforced predictable components on decadal time scales, distinguish these components from forced predictable components, and to assess the reliability of model predictions of these components. The question of whether anthropogenic forcing changes decadal predictability, or gives rise to new forms of decadal predictability, also will be investigated.

2 Diagnosis of Decadal Predictability

Our work has established a scientific basis for decadal predictions by explicitly identifying patterns in climate models that are predictable on decadal time scales and showing that these structures also are predictable in the observed climate system. Our research clearly demonstrates that certain patterns in climate models are predictable on decadal time scales. Specifically, we identified patterns that maximize Average Predictability Time in multiple pre-industrial control runs. Control runs have no interannual variations in climate forcing, hence the multidecadal variability found in such simulations are generated internally by the climate system irrespective of anthropogenic or natural forcing. The single most predictable component in CMIP5 models is shown in fig. 1a, which has loadings concentrated in the North Atlantic and North Pacific, consistent with previous studies. This pattern is discussed in more detail in Jia and DelSole (2012a), and is similar to the most predictable pattern in CMIP3 models (DelSole et al., 2011). The APT of this component is well separated from those of the other components and statistically significant in each model individually. The skill of predicting this component in pre-industrial control runs, using a linear regression model derived from independent control simulations, is shown in fig. 1b. As indicated in the figure, different models show widely different levels of predictability, with some models showing insignificant predictability after two years while others showing significant predictability as long as 10 years. Thus, models agree that certain SST patterns are predictable on decadal time scales, but disagree on the magnitude of that predictability. The time series produced by projecting the component onto observational estimates of sea surface temperature from the Extended Reconstruction data set (ERSSTv3) is shown in fig. 1c. The figure shows that the time series fluctuates strongly on multidecadal time scales. This result is noteworthy because there is no guarantee that the most predictable component in climate models is predictable (or slowly varying) in observations. Fig. 1c also shows the observed Atlantic Multidecadal Oscillation (AMO) index. The strong correlation between the two time series confirms that the most predictable component in the CMIP5 models projects on a well documented component of multi-decadal variability in observations. **The fact that this component can be predicted on decadal time scales in some models provides a scientific basis for decadal prediction of surface air temperature**.

In total, we identify four robust SST patterns that are predictable on decadal time scales in climate models. The second and fourth patterns correspond to the Pacific Decadal Oscillation and ENSO, respectively, while the third pattern appears to be a mixture of the above patterns (not shown). In addition, a generalized version of APT analysis, in which a linear regression model is used to predict surface air temperature or precipitation *over land*, using SST as a predictor, was used to show that temperature is predictable for 3-6 yrs and that precipitation is predictable for 1-3 yrs, depending on continent (Jia and DelSole, 2011). Results of this research have been summarized in the invited review paper DelSole et al. (2015a).

The most predictable component of annual mean precipitation in the CMIP5 models is shown in fig. 2. In contrast to surface air temperature discussed above, the most predictable precipitation component has maximum loadings in the tropical pacific, and the models more or less agree that the predictability persists only for two or three years. Because this component has been statistically optimized, **these results demonstrate that predictability of annual mean precipitation is at most three years, and hence decadal predictability of precipitation is weak or non-existent in the CMIP5 models.**

3 Optimal Determination of Time-Varying Climate Change Signals

We developed an optimization procedure that identifies components that maximize detectability of forced responses. The resulting components are called *most detectable components* and discussed in Jia and DelSole (2012b). The most detectable component of near-surface air temperature in CMIP5 models is shown in fig. 3a. The pattern corresponds to temperature changes in the same direction nearly everywhere. Combining this fact with the increasing trend in each time series implies that the bulk of the detectability is due to twentieth century warming. Note that the loadings are relatively uniform in the tropics, but are

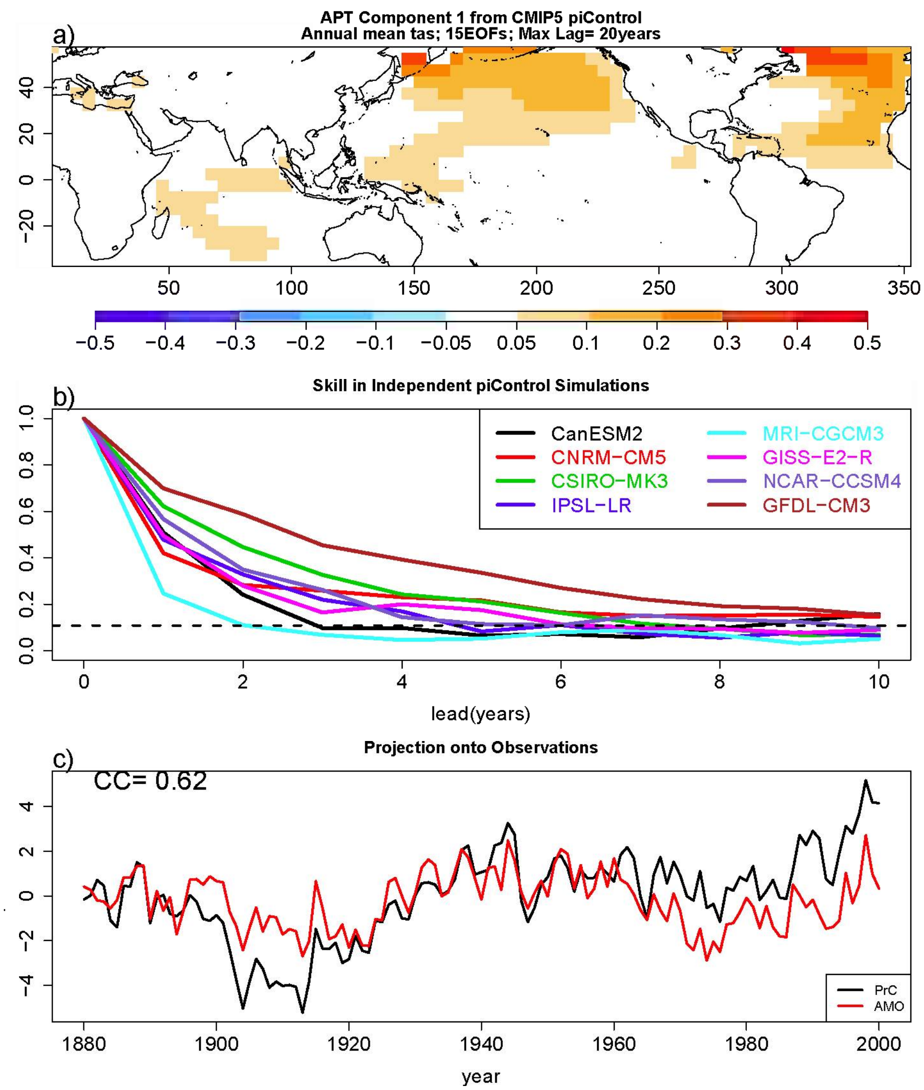

Figure 1: a) The most predictable component of annual mean near-surface air temperature over the ocean between $40°$S and $60°$N derived from eight CMIP5 control simulations; b) the squared multiple correlation of the most predictable component in each model; c) the (scaled) AMO index (red), and the time series produced by projecting the pattern onto observations (black). The dashed line in b) indicates the 5% significance level. The spatial structure shows the change in temperature (in degree kelvin per unit variate) relative to the climatological mean. The correlation coefficient between the two time series in c) is indicated in the top left of the panel.

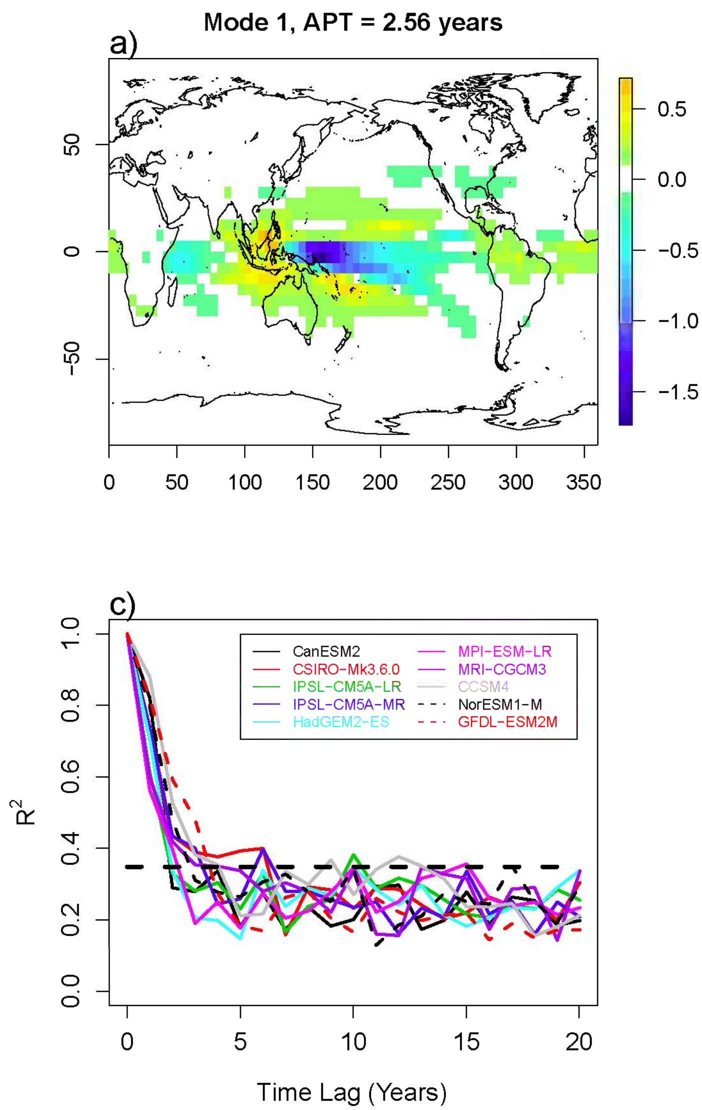

Figure 2: Most predictable component of annual mean precipitation derived from the pre-industrial control runs of the CMIP5 data set (top), and the squared correlation skill of this component in individual models when predicted from a linear regression model using forty principal components as predictors.

very weak in the North Atlantic, in contrast to the most predictable component shown in fig. 1. These differences in spatial structure provide the basis for separating forced and unforced variability in observations (DelSole et al., 2011). The detectability of the pattern, measured by the forced-to-unforced variance ratio, is shown in fig. 3c. The leading component is detectable in all models (i.e., all values for the first component lie above the upper significance threshold). In contrast, the second is not detectable in some models, and the degree of detectability is model dependent: the ratio of the largest to smallest variance ratio for the first component exceeds ten. Time series of the most detectable component in independent twentieth century runs are shown in fig. 3b. Each time series shows a clear trend, interrupted by sudden coolings coincident with major volcanic eruptions. The variability due to the trend greatly exceeds the variability of the pre-industrial control simulations (indicated by the error bar), demonstrating that this component is significantly detectable in the models. Note that the model ensembles become separated at the end of the twentieth century, indicating that a large fraction of the uncertainty in forced response comes from model differences.

Jia and DelSole (2012b) applied the above optimization procedure to determine the most detectable pattern in the CMIP3 models of precipitation over land. No single pattern of five-year mean JAS precipitation over land could be detected in all the models, indicating that **the forced response of precipitation is not robust across state-of-the-art climate models**. Because the procedure is optimal, we can claim that *no* five-year mean precipitation pattern is *consistently* detectable in models. It should be recognized, however, that this conclusion does not preclude detectability of other expressions of the signal, such as ones constructed from other time averages, statistics, or time-lag information. The amplitude of the climate change statistic in different twentieth century simulations is shown in fig. 4. The figure indicates that while most models show no detectability of precipitation (as evidenced by the fact that their time series lie within the 90% confidence interval for unforced variability), the models that do show detectability of five-year mean JAS precipitation exhibit two very different characters: a systematic trend, and enhanced frequency of extreme values. These results appear to be a significant advancement in detection capabilities, as previous attempts to identify forced precipitation patterns were restricted to zonal averages within certain latitudinal bands.

4 Empirical Decadal Prediction

The above studies demonstrate that temperature and precipitation are predictable on multi-year time scales *in the models*, but they do not show that this model-derived predictability translates into skillful predictions of observed anomalies. To test whether these components are predictable in observations, DelSole et al. (2013) constructed a linear regression model from the pre-industrial control runs, and used this model to predict *observations*. In this new forecast system, the forced component is predicted using the most detectable component shown in fig. 3, with the amplitude determined by a least squares fit of this pattern for each year to the ensemble mean, multi-model mean of eight CMIP5 twentieth century simulations. These simulations were initialized at some random point from a pre-industrial simulation and

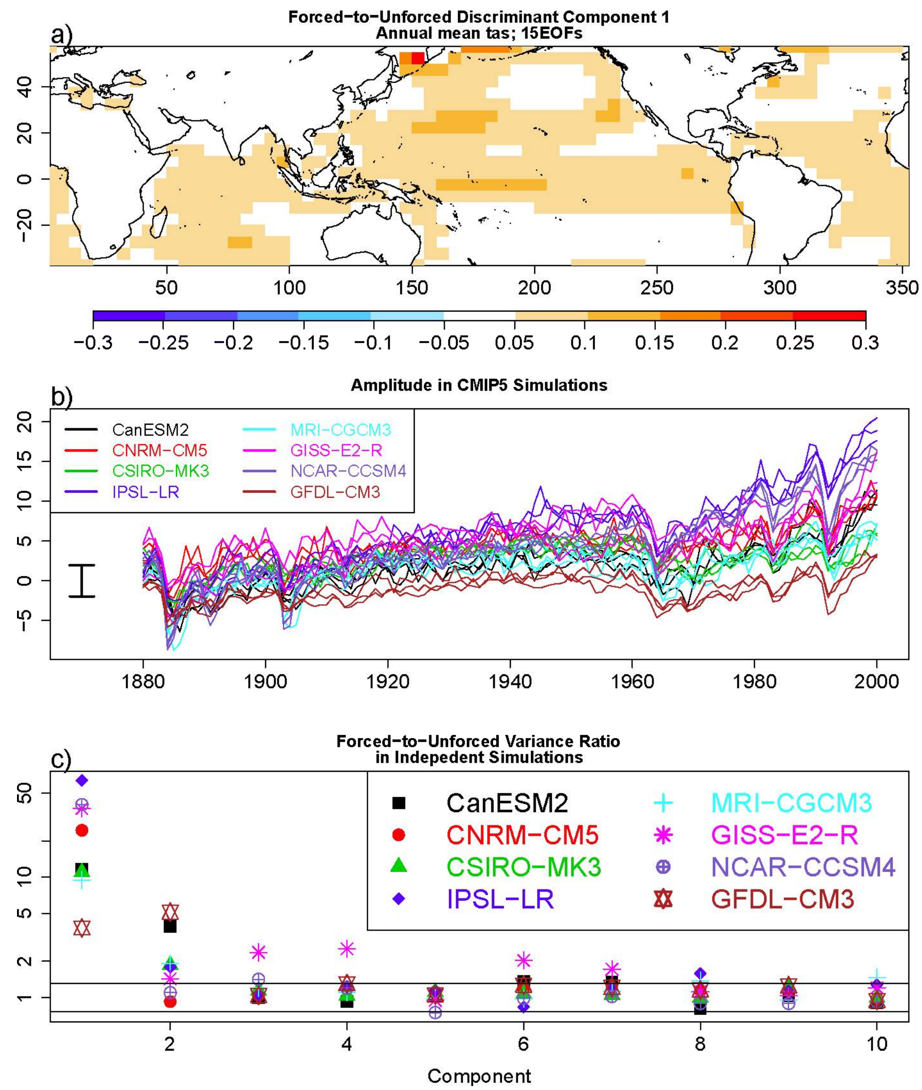

Figure 3: a) The most detectable component, on average, of annual mean near-surface air temperature over the ocean between $40°$S and $60°$N derived from eight CMIP5 control simulations; b) the projection of the most detectable pattern on independent twentieth century simulations; c) the forced-to-unforced variance ratios of each discriminant component in each model. The error bar in b) indicates the 2.5 and 97.5 percentiles in the pre-industrial control simulations. For plotting b), each model time series has been normalized by the standard deviation of the pre-industrial control runs of the respective model, and ensemble members from the same model have the same color. The two lines in c) show the 2% confidence interval of the variance ratio under the null hypothesis of equal variance. The spatial pattern has units of degree Kelvin per unit variate relative to the pre-industrial mean.

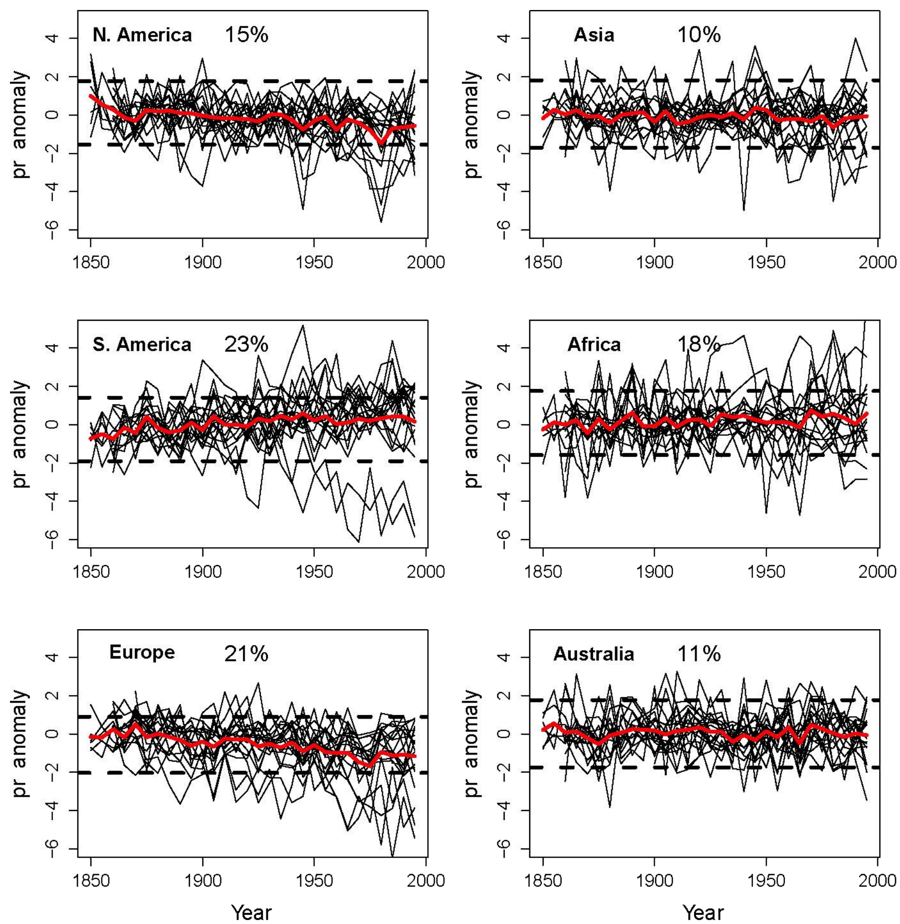

Figure 4: Time series of the most detectable component, on average, of JAS mean precipitation over six continents. Individual twentieth century runs are shown as thin black lines, and the ensemble mean is shown as the thick red curve. The time series were calculated from 19 independent twentieth century runs. The horizontal dashed lines indicate the 5th and 95th percentiles estimated from independent control runs. The percentages in each panel gives the percent of twentieth century realizations that fall outside the 5th and 95th percentiles of the control run. The pre-industrial control mean of each respective model has been removed from the time series.

then forced with anthropogenic and natural forcing characteristic of the twentieth century. As a result, all prediction skill from these simulations come from the forced response, not from observational initial condition information. The skill of the model-predicted forced component therefore provides a measure of how well the forced component is predicted by dynamical models. To predict the unforced component, we use a multivariate regression model that predicts the mean sea surface temperature in a given year based on the state of this field at an antecedent year. We avoid complex validation issues by constructing a single multivariate regression model trained on pre-industrial control simulations, which affords a much larger sample size is available for estimation, and also leaves the entire observational record to serve as independent verification data. On the other hand, a regression model trained on numerical simulations (instead of observations) may be unable to capture realistic SST variability. In a sense, the regression model represents a simplified version of the dynamical models from which it was trained, and hence provides a way to validate dynamical models without performing forecasts with the dynamical models themselves. To mitigate sensitivity to the choice of dynamical model, a single regression model is trained on control runs from eight different climate models. The resulting prediction model is exactly the same model used to derive the most detectable component in fig. 1.

To quantify prediction skill, we select three indices, namely the area-weighted average over the North Pacific, North Atlantic, and globe, and calculate one minus the ratio of the mean square prediction error to the climatological variance of these indices. The skill of the forced response estimate in predicting 20C observations is shown in fig. 5 as the blue line. The sum of forced and unforced temperature anomalies is predicted by adding the estimated forced response to the regression model prediction. The skill of the resulting predictions is shown as the black curve in fig. 5. The skill is positive at all lead times. However, the skill of the forced plus unforced prediction approaches the skill of the forced component alone. On the other hand, the skill generally exceeds the skill of a persistence forecast, demonstrating that the regression model "adds value" beyond mere persistence. The skill is largest for the global domain, reflecting the dominance of forced predictability on global space scales. The fact that the skill of the combined forced-unforced prediction is larger than the skill of the forced response prediction, especially in the first few lead years, demonstrates that **initial condition information contributes to skill on multi-year time scales.**

The above forecast system can skillfully predict SSTs on multi-year time scales, but it combines different patterns together so it is unclear whether its skill comes from the predictable patterns discussed in previous sections. To clarify this point, we show in fig. 6 the skill of our forecast system in predicting the first four predictable components. The skill is calculated by first forecasting the entire SST field, then projecting each predictable component onto the forecast field and observations, then calculating the skill of the resulting indices. The figure shows that the skill of each predictable component exceeds the skill of the forced component alone, and exceeds the skill of a persistence forecast. In the case of the third predictable component, the skill of a persistence forecast is nearly the same as that of our model, implying that the skill of this component comes primarily from persistence. In contrast, our forecast model predicts the other components with more skill than a persistence

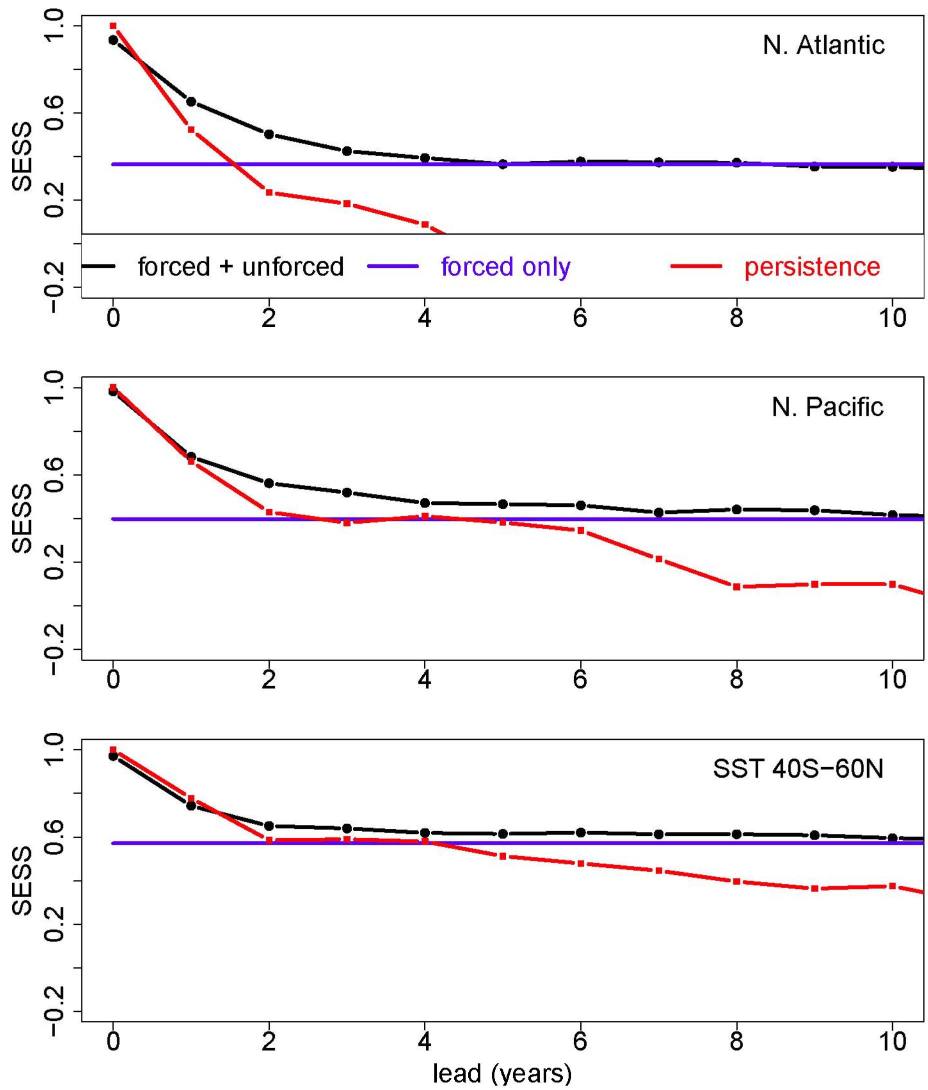

Figure 5: The skill of predicting observed annual mean near-surface temperatures over the ocean in the North Atlantic (top), North Pacific (middle), and between 40°S and 60°N (bottom) during 1910-2000, using an estimate of the forced component derived from twentieth-century simulations from eight CMIP5 models (blue line), and using this estimate of the forced response plus a multivariate regression model to predict unforced variability (black curve). The skill of a persistence forecast also is shown (red curve). Skill is measured by the squared error skill score (SESS), defined as one minus mean square error divided by the observed climatological variance.

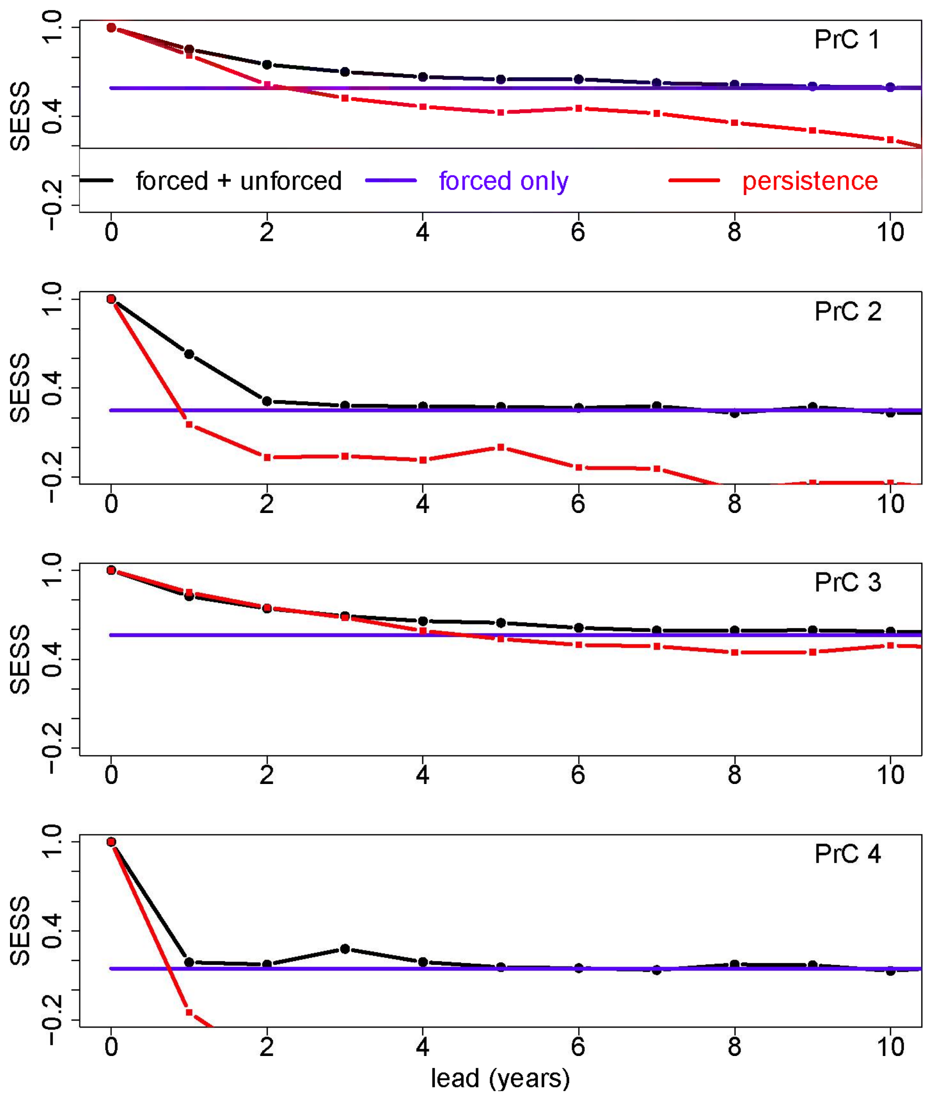

Figure 6: Same as fig. 5, except for predicting the first four predictable components obtained by maximizing APT.

forecast. In fact, for the second and fourth predictable components, the skill of a persistence forecast is *negative*. This negative skill is consistent with our hypothesis that the predictability in the Pacific comes from the *interaction* of components rather than persistence. These results are summarized in DelSole et al. (2015a).

5 Changes in Internal Variability Due to Anthropogenic Forcing

We also investigated how the variability of temperature will change due to global warming. There is growing evidence that the frequency and intensity of heat extremes will increase in response to increasing greenhouse gas concentrations. However, there is considerable disagreement as to whether these changes can be explained as a simple shift of the probability distribution of temperature, or whether the shape of the distribution (especially the variance) also is changing. LaJoie and DelSole (2015) quantified changes to internal variability of seasonal- and annual-mean 2m temperature in response to anthropogenic forcing using climate models driven by twenty-first century high emissions scenarios. The statistical significance of the spatially distributed changes to variance is tested using a new multivariate technique. All climate models examined predict significant changes to internal variability of temperature in response to anthropogenic forcing. The models consistently predict decreases to temperature variance in regions of seasonal sea-ice formation and across the Southern Ocean during the twenty-first century. While more than half the models predict significant changes in variance over ENSO regions and in the North Atlantic Ocean, the direction of this change is model dependent. Seasonal mean changes are remarkably similar to annual mean changes, but there are some model-dependent exceptions. Some models predict future variability that is more than double their preindustrial variability, raising questions about the adequacy of doubling uncertainty estimates to test robustness in detection and attribution studies. The results show that **climate models agree that temperature variability will decrease over areas influenced by changing sea-ice, but they disagree about changes in other areas of the globe, especially the tropical Pacific, which is the center for El Nino phenomena**.

6 Improved Estimates of Aerosol Cooling

We have developed a new framework for quantifying the detectability of anthropogenic aerosols. This new framework allows one to systematically explore a variety of variables and spatio-temporal details to maximize detection of anthropogenic aerosol forcing. Applying this methodology to CMIP5 simulations shows clearly that a combination of temperature and precipitation provides the maximum detectability of anthropogenic aerosols in a perfect model world (Yan et al., 2015). Surprisingly, the detectability does not require any spatial gradient information (i.e., aerosol forcing can be detected using only global mean temperature and precipitation) (DelSole et al., 2015b). Unfortunately, applying this approach to real observations is problematic because the errors in precipitation observations are large; for instance, different observational estimates of global-mean, annual-mean precipitation are not significantly correlated with each other (DelSole et al., 2015b).

7 Contribution of Individual Collaborators

Timothy DelSole is the lead PI for this project and is responsible for all aspects of the proposed research. He supervised research conducted by the PhD graduate students Emerson LaJoie, Xiaoxin Yan, and Liwei Jia (now graduated). Dr. Jia performed the analysis leading to Jia and DelSole (2012a), Jia and DelSole (2012b), DelSole et al. (2013). Emerson LaJoie performed the analysis leading to LaJoie and DelSole (2015). Xiaoxin Yan entered the PhD program in Fall 2011 and performed the analysis leading to Yan et al. (2015) and DelSole et al. (2015b). Dr. Jia has been working at GFDL since 2013. Michael Tippett has collaborated on all aspects of the proposed research.

8 Publications

The following papers generated as part of this research project have been published in peer-reviewed journals or books:

- Jia and DelSole (2012a)
- Jia and DelSole (2012b)
- DelSole et al. (2013)
- DelSole et al. (2015a)
- LaJoie and DelSole (2015)
- Yan et al. (2015)
- DelSole et al. (2015b)

References

DelSole, T., L. Jia, and M. K. Tippett, 2013: Decadal prediction of observed and simulated sea surface temperatures. *Geophys. Res. Lett.*, **40**, 2773–2778.

DelSole, T., M. K. Tippett, and L. Jia, 2015a: Multi-year prediction and predictability. *Climate Change: Multidecadal and Beyond*, C.-P. Change, M. Ghil, M. Latif, and J. M. Wallace, Eds., World Scientific Publishing, World Scientific Series on Asia-Pacific Weather and Climate, Vol. 6, chap. 14, 219–233.

DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Climate*, **24**, 909–926.

DelSole, T., X. Yan, and M. K. Tippett, 2015b: Inferring aerosol cooling from hydrological sensitivity. *J. Climate*, **submitted**.

Jia, L. and T. DelSole, 2011: Diagnosis of multiyear predictability on continental scales. *J. Climate*, **24**, 5108–5124.

Jia, L. and T. DelSole, 2012a: Multi-year predictability of temperature and precipitation identified in climate models. *Geophys. Res. Lett.*, **39**, doi:10.1029/2012GL052 778.

Jia, L. and T. DelSole, 2012b: Optimal determination of time-varying climate change signals. *J. Climate*, **25**, 7122–7137.

LaJoie, E. and T. DelSole, 2015: Changes to internal variability in response to anthropogenic forcing: A new field significance test. *J. Climate*, **submitted**.

Yan, X., T. DelSole, and M. K. Tippett, 2015: What surface observations are important for separating the influences of anthropogenic aerorsols from other forcings? *J. Climate*, **submitted**.