Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts

Hye-Mi Kim,[1] Peter J. Webster,[1] and Judith A. Curry[1]

[1] This study assesses the CMIP5 decadal hindcast/forecast simulations of seven state-of-the-art ocean-atmosphere coupled models. Each decadal prediction consists of simulations over a 10 year period each of which are initialized every five years from climate states of 1960/1961 to 2005/2006. Most of the models overestimate trends, whereby the models predict less warming or even cooling in the earlier decades compared to observations and too much warming in recent decades. All models show high prediction skill for surface temperature over the Indian, North Atlantic and western Pacific Oceans where the externally forced component and low-frequency climate variability is dominant. However, low prediction skill is found over the equatorial and North Pacific Ocean. The Atlantic Multidecadal Oscillation (AMO) index is predicted in most of the models with significant skill, while the Pacific Decadal Oscillation (PDO) index shows relatively low predictive skill. The multi-model ensemble has in general better-forecast quality than the single-model systems for global mean surface temperature, AMO and PDO. **Citation:** Kim, H.-M., P. J. Webster, and J. A. Curry (2012), Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts, *Geophys. Res. Lett.*, *39*, L10701, doi:10.1029/2012GL051644.

1. Introduction

[2] The prediction of decadal climate variability against a background of global warming is one of the most important and challenging tasks in climate science. Not only does natural variability have a large-amplitude influence over broad regions of the globe, it is an integral component of climate variability that modulates low-frequency climate phenomena as well as extreme climate events such as tropical cyclone activity. On decadal timescales, some aspects of internal climate variability may be predictable [*Collins and Allen*, 2002; *Smith et al.*, 2007; *Keenlyside et al.*, 2008; *Meehl et al.*, 2009, 2010; *Pohlmann et al.*, 2009; *Mochizuki et al.*, 2012]. However, the actual prediction skill of natural climate variability on decadal timescales using various current climate models has received little attention [*van Oldenborgh et al.*, 2012].

[3] The Coupled Model Intercomparison Project Phase 5 (CMIP5) has devised an innovative experimental design to assess the predictability and prediction skill on decadal time scales of state-of-the-art climate models, in support of the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report [*Taylor et al.*, 2012]. The decadal predictability and prediction skill of individual models have been analyzed separately for multi-year prediction horizons over different time periods and regions [*Pohlmann et al.*, 2009; *Fyfe et al.*, 2011; *Chikamoto et al.*, 2012; *Mochizuki et al.*, 2012]. However, the CMIP5 decadal predictions from different models have not been evaluated and compared using the same evaluation matrix. The choice of one model over the other, or the use of sets of models in a multi-model ensemble (MME), requires information that compares the predictions of individual models. Here, we compare the ability of currently available CMIP5 decadal hindcasts to simulate the mean climate and decadal climate variability from individual coupled models and a multi-model ensemble. We focus on the surface temperature and two dominant internal climate modes: the Atlantic Multidecadal Oscillation (AMO) and Pacific Decadal Oscillation (PDO). This study addresses how well the CMIP5 multi-model decadal hindcasts simulate the spatio-temporal climate variability.

2. Data and Models

[4] This study compares the CMIP5 decadal hindcasts and forecasts conducted by seven modeling centers. Each decadal prediction includes at least 3 and as many as 10 ensemble members generated by slightly different initial conditions. The data consists of simulations over 10 year period that are initialized every five years during the period 1960/1961 to 2005/2006 [*Taylor et al.*, 2012]. A brief summary of each model's experimental configuration is presented in Table S1 (see Text S1 in the auxiliary material).[1] The annual mean refers to the average from January to December for each year.

[5] Surface temperature data from the ERA40 [*Uppala et al.*, 2005] before 1979 and from the ERA Interim [*Berrisford et al.*, 2009] after 1979 are used to evaluate the predictions. The Extended Reconstructed Sea Surface Temperature Version 3 (ERSST.v3b) [*Smith et al.*, 2008] is used to define the PDO and AMO indices. All data from model hindcasts and observations are interpolated to horizontal resolution as $2.8125°$ longitude and latitude. For the observations, the long-term mean is removed by subtracting climatological means for the entire period from 1960 to 2010. The model forecast anomaly is calculated as $Y'_{j\tau} = Y_{j\tau} - \bar{Y}_\tau$, where Y is the ensemble-average prediction, Y' is the anomaly of the raw forecast with respect to the forecast

---

[1]School of Earth and Atmospheric Science, Georgia Institute of Technology, Atlanta, Georgia, USA.

Corresponding author: H. Kim, School of Earth and Atmospheric Science, Georgia Institute of Technology, 311 Ferst Dr., Atlanta, GA 30332, USA. (hyemi.kim@eas.gatech.edu)
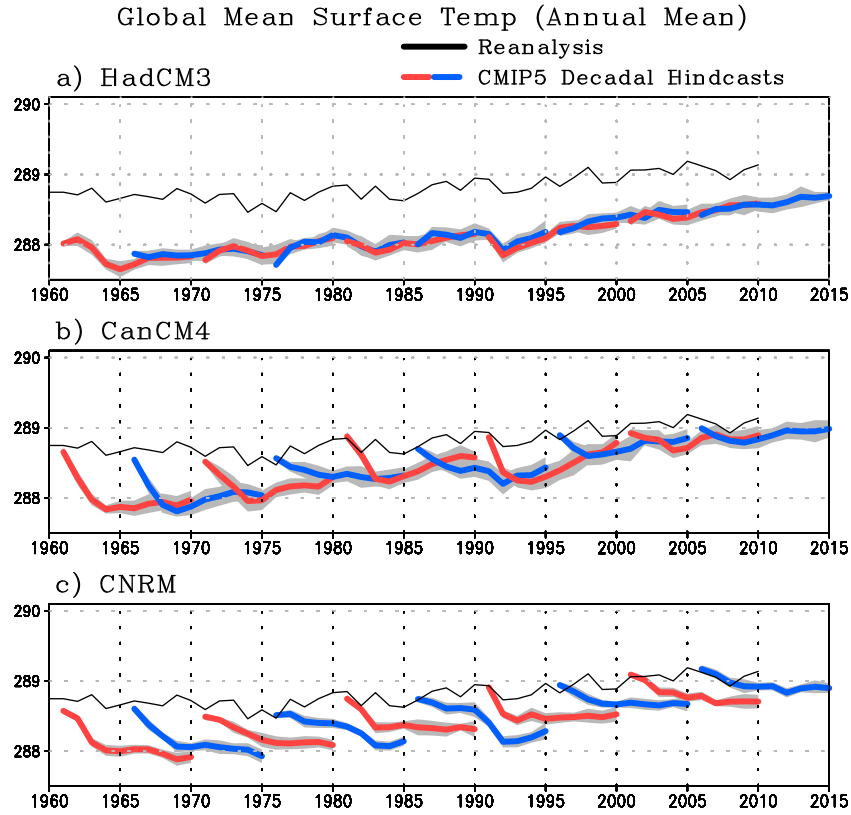
Figure 1. Time series of globally averaged annual-mean surface temperature [K] for reanalysis (black) and the ensemble-mean of the CMIP5 decadal hindcasts and forecasts (red and blue) for (a) HadCM3, (b) CanCM4, (c) CNRM, (d) MIROC4h, (e) MIROC5, (f) MRI and (g) CFSv2. Gray shades represent the ranges of one standard deviation of the ensembles in each hindcasts.

average $\bar{Y}$, j is the starting year (n = 1, 2, …, 10) and τ is the forecast lead year. $\bar{Y}$ is calculated as $\bar{Y}_\tau = \frac{1}{n}\sum_{j=1}^{n} Y_{j\tau}$ in the period only when observational data is available. The equally weighted average from total 52 ensemble members of seven hindcast experiments provides the values for the MME.

3. Prediction Skill Assessment for CMIP5 Decadal Hindcasts

3.1. Global Surface Temperature

[6] The model prediction skill is examined by comparing the annual mean surface temperature from the observation and hindcasts of each model. Figure 1 shows the evolution of the annual global surface temperature ($Y_{j\tau}$) from the reanalysis from 1960 to 2010 and ensemble mean of each model 10 year hindcast/forecast from 1961 to 2015 (initialized every 5 years). Most of the models simulate lower global mean surface temperature than the reanalysis during the entire period, except MIROC4h (Figure 1d), MIROC5 (Figure 1e) and CFSv2 (Figure 1g). Several models, including CanCM4 (Figure 1b) and CNRM (Figure 1c), which are initialized close to the observed state (full field initialization), drift towards the model climate. Models that are initialized with anomaly assimilation (anomaly initialization) also show climate drift during predictions (MIROC4h, MIROC5 and MRI: Figures 1d–1f, respectively). Most of the models appear to overestimate the 1960–2010

trend in their hindcasts. Removing the average along the actual time ($Y'_{j\tau}$) leads to a skewed outcome, whereby the MME (Figure S1 in Text S1) and each models (not shown) predict less warming or even cooling compared to the reanalysis in the earlier decades, and too much warming in recent decades. Figure 2 shows the observed and predicted trend (slope in a linear regression) in $Y'_{j\tau}$ as a function of forecast lead time (τ) after applying a four-year average to filter out the high frequency variability. The observed trend is calculated in the same manner of the trend in hindcasts. The systematic overestimation in the trend throughout the integration period is obvious in all hindcasts except CFSv2.

[7] To examine the prediction skill of the individual models and also the MME, the average for the lead times of 1-year, 2–5 year and 6–9 year mean surface temperature anomaly field ($Y'_{j\tau}$) is compared with reanalysis (Figure 3 and Figure S2 in Text S1). We measure the prediction skill in terms of the anomaly correlation coefficients:

$$CORR(\tau) = \frac{\sum_{j=1}^{n}\left[Y_{j\tau} - \bar{Y}_\tau\right]\left[O_{j\tau} - \bar{O}\right]}{\sqrt{\sum_{j=1}^{n}\left[Y_{j\tau} - \bar{Y}_\tau\right]^2 \sum_{j=1}^{n}\left[O_{j\tau} - \bar{O}\right]^2}},$$

where the O's are the observed field. The correlation coefficients are calculated over the ensemble mean for the hindcasts of each models. The results are almost the same if the observed anomaly is calculated based on the climatology $\bar{O}_\tau$.
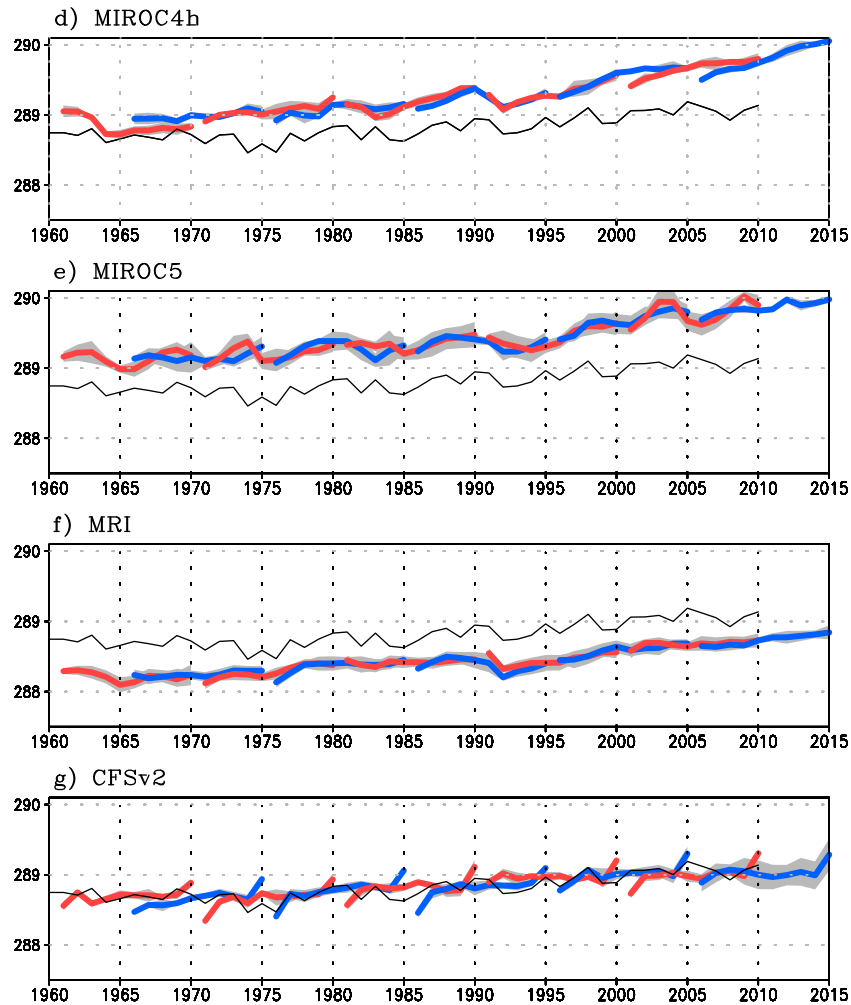
Figure 1. (continued)

[8] All models show high skill (greater than the 95% confidence level) in forecasting surface temperature anomalies over the Indian, North Atlantic and the western Pacific Oceans up to 6–9 years (Figure 3 and Figure S2 in Text S1). However, the equatorial Pacific and North Pacific Ocean regions show less prediction skill after 2–5 years (Figure 3). For 6–9 years, the predictive skill does not change much compared to 2–5 years (Figure S2 in Text S1). The relatively long prediction skill appears over the region where the externally forced component and low-frequency climate variability is dominant [*Keenlyside et al.*, 2008; *Meehl et al.*, 2009; *Pohlmann et al.*, 2009; *Chikamoto et al.*, 2012, *Mochizuki et al.*, 2012; *Oldenborgh et al.*, 2012]. High prediction skill also occurs in the tropical Atlantic SST which is an important factor in climate variability in that region and beyond [*Keenlyside et al.*, 2008]. Comparing the globally averaged skill for each model's hindcasts shows the highest skill occurring for the MME over the entire period (not shown). The relatively low skill in MIROC4h and CFSv2 is possibly due to the smaller number of ensemble members (i.e. 3 and 4 members, respectively) compared to other models (i.e. 6–10 members) as larger number of

ensemble members generally results in representing a higher skill in the ensemble mean.

3.2. Decadal Climate Variability

[9] To examine the prediction skill of natural internal modes of climate variability, the simulation of AMO and PDO indices is compared with observations. The AMO and PDO are the dominant decadal oscillations over the North Atlantic Ocean [*Schlesinger and Ramankutty* 1994; *Enfield et al.*, 2001] and North Pacific Ocean [*Mantua et al.*, 1997], respectively, and are the most predictable components of internal climate variability [e.g., *Keenlyside et al.*, 2008; *Mochizuki et al.*, 2010]. The AMO index is defined as the area averaged annual mean sea surface temperature (SST) anomaly averaged over the North Atlantic from 80°W to 20°W and from 0° to 70°N for both the simulations and observation. The simulated PDO index is defined as the normalized time series based upon projections of predicted annual mean SST anomaly in the North Pacific Ocean poleward of 20°N, onto the leading EOF spatial pattern from the observed annual mean SST anomaly. Both in the observations and model hindcasts, the SST anomalies are
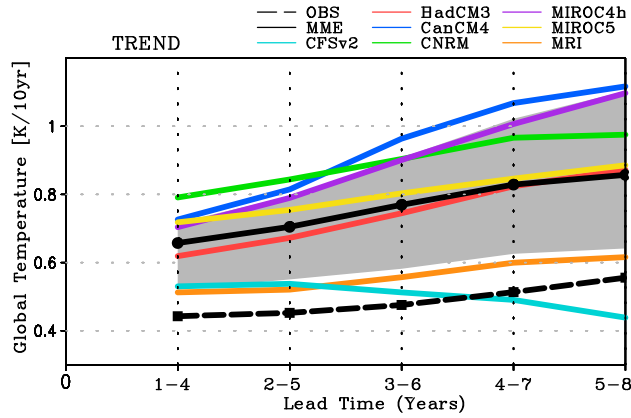
Figure 2. Trend [K/10yr] for the global mean annual temperature anomaly predicted by MME and ensemble-mean of each CMIP5 decadal hindcasts as a function of lead time. Black dashed line represents the trend in the observation. Gray shades represent the ranges of one standard deviation of the ensemble-mean in each hindcasts.

detrended before calculating these indices to remove the externally-forced variation [*Oldenborgh et al.*, 2012]. A four-year running average is applied to both indices to filter out higher interannual frequencies. Figure S3 in Text S1 shows the variation of the AMO and PDO indices from observations and the MME hindcast. Both indices show strong decadal variability. The gray shades in Figure S3 represent the ranges of one standard deviation of the ensemble mean in each hindcast.

[10] The predictive skill for the AMO and PDO index is measured by correlation coefficient and root-mean-square error (RMSE) between the simulations and observation. Figure 4 shows the correlation coefficient as a function of lead-time for the MME and the ensemble mean of individual models. For representing confidence limits of significance, the correlations and RMSE of the persistence prediction are included (Figure 4 and Figure S4 in Text S1). Horizontal lines in each figure represent the confidence level (Figure 4) and observed standard deviation (Figure S4 in Text S1), respectively. For the AMO prediction, the correlation coefficients and RMSE of almost all models represent significant
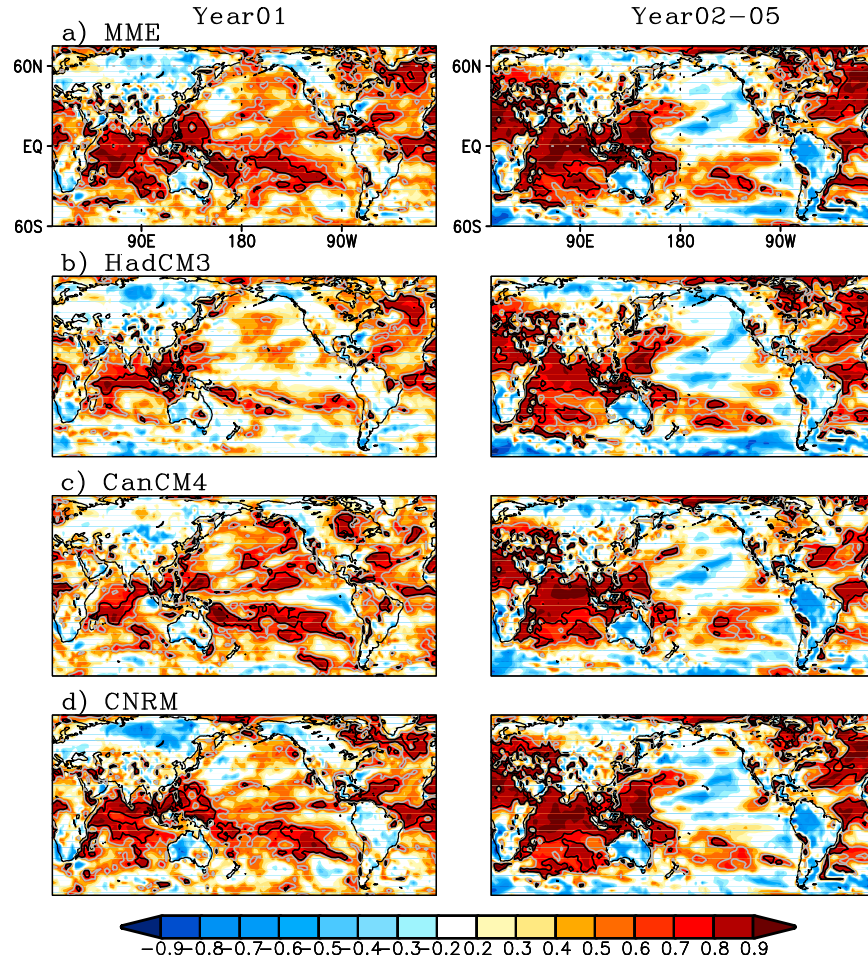

Figure 3. The spatial distribution of temporal correlation coefficients for the annual mean surface temperature anomaly between reanalysis and decadal hindcasts at forecast years 1 and 2–5 years average. The values show the correlation coefficients from ensemble-mean for each of models for (a) MME (b) HadCM3, (c) CanCM4, (d) CNRM, (e) MIROC4h, (f) MIROC5, (g) MRI and (h) CFSv2. Solid black (gray) line represents statistical significance of the correlation coefficients at 99% (95%) confidence level.
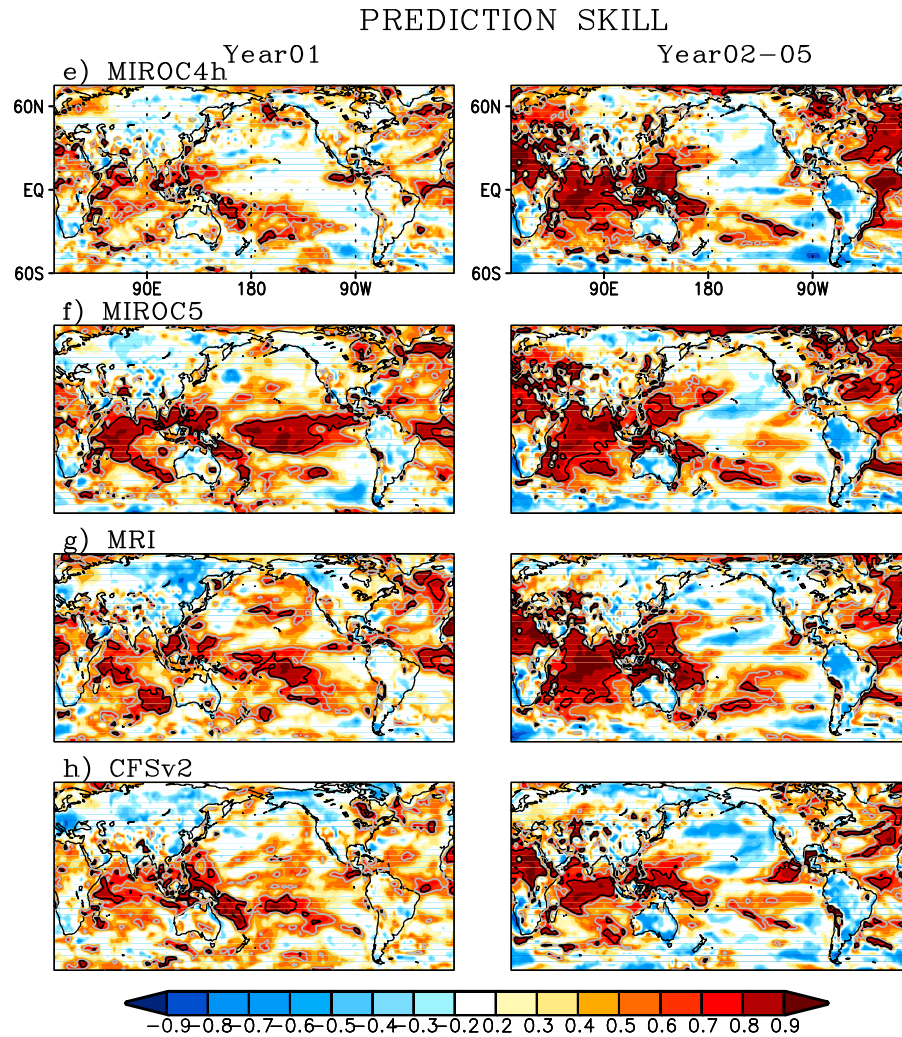
Figure 3. (continued)

skills (Figure 4a and Figure S4a in Text S1). After 1–4 years, the MME, HadCM3, CNRM and MIROC4h show greater skill than the persistence prediction. After 3–6 years, most of the models have greater and significant skill than the persistence prediction (high correlation than persistence and smaller RMSE than the persistence and observed amplitude). The MME represents more skillful results than most of the individual model predictions over the entire prediction period.

[11] The prediction skill in PDO index is lower than AMO, in agreement with recent studies [*Oldenborgh et al.*, 2012]. The correlation coefficient of the PDO index shows predictive skill over 90% confidence level in MME and CanCM4 for 1–4 and 2–5 years. CanCM4 remains being above 90% during 3–6 years and MIROC5 is far above 95% for 3–6 years. The MME shows a decrease in skill for lead times beyond 3–6 years (Figure 4b). The correlation coefficients of almost all models represent insignificant skills for the PDO index over the entire period. The correlation coefficient is less than the persistence prediction and the errors of all models are larger than the observed PDO

amplitude. The MME shows more skillful results than most of the individual model predictions.

4. Conclusion

[12] We have assessed the CMIP5 decadal hindcast/forecast simulation performance of seven state-of-the-art ocean-atmosphere coupled models. Most of the models produce cooler than observed global mean temperature during the entire period and overestimate the observed trend in their hindcasts. All models show high prediction skill for surface temperature up to 6–9 years over the Indian Ocean, the North Atlantic and the western Pacific Oceans, while showing lower predictive skill over the equatorial Pacific and North Pacific Ocean. The AMO index is relatively well predicted in all models for the entire prediction period with a significant skill, while the predictive skill for the PDO index is relatively low for the entire period.

[13] Although the MME does not outperform all of the constituent models for every forecast skill metric, it has in general better forecast quality than the single models for
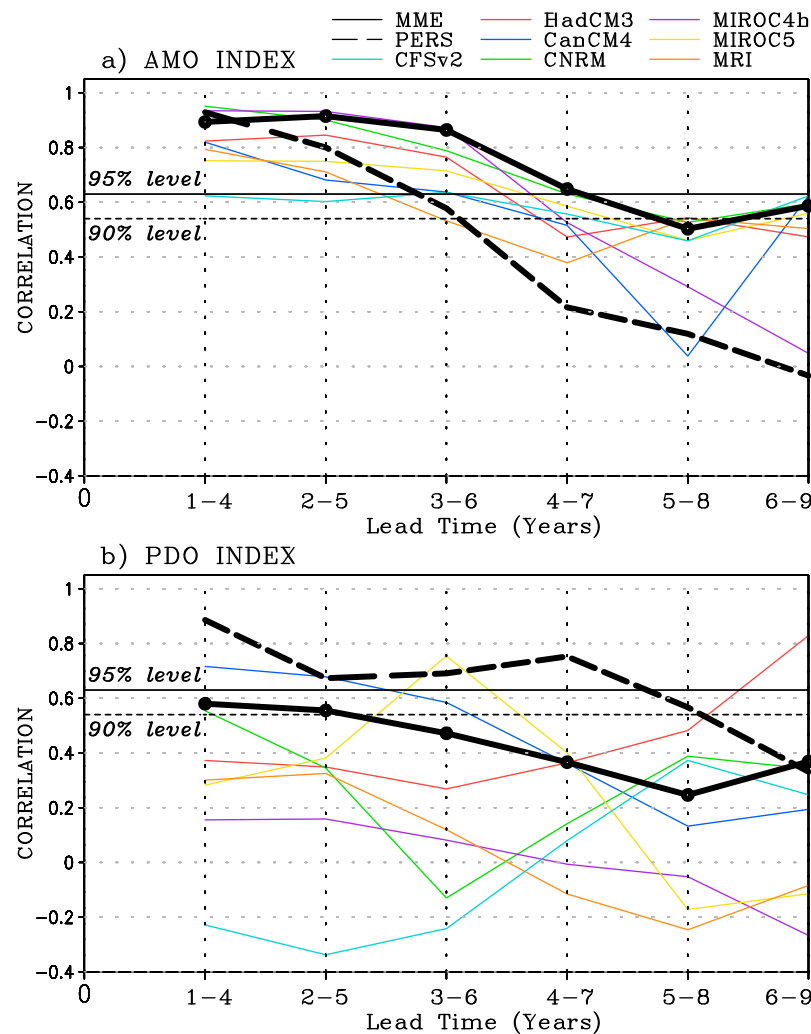
Figure 4. Correlation coefficients for the (a) AMO and (b) PDO index predicted by MME, persistence (PERS) and ensemble-mean of each CMIP5 decadal hindcasts as a function of lead time (years). Solid (dashed) horizontal line represents statistical significance of the correlation coefficients at 95% (90%) confidence level.

global mean temperature, AMO and PDO. This study partly supports the utility of the multi-model ensemble approach in overcoming the systematic model biases from individual models and in enhancing decadal predictability. It should be noted that not all modeling centers have thus far released their decadal predictions for CMIP5. Additional intercomparison will be conducted when the other CMIP5 simulations are made available.

[15]   The Editor thanks the two anonymous reviewers for their assistance in evaluating this paper.

References

Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala (2009), The ERA-Interim archive, *ERA Rep. Ser. 1*, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Chikamoto, T., et al. (2012), Predictability of a stepwise event of Pacific climate change during the late 1990s in hindcast experiments by MIROC, *J. Meteorol. Soc. Jpn.*, *90A*, 1–21.

Collins, M., and M. R. Allen (2002), Assessing the relative roles of initial and boundary conditions in interannual to decadal climate prediction, *J. Clim.*, *15*, 3104–3109, doi:10.1175/1520-0442(2002) 015<3104:ATRROI>2.0.CO;2.

Enfield, D. B., A. M. Mestas Nunez, and P. J. Trimble (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S, *Geophys. Res. Lett.*, *28*, 2077–2080, doi:10.1029/ 2000GL012745.

Fyfe, J. C., W. J. Merryfield, V. Kharin, G. J. Boer, W.-S. Lee, and K. von Salzen (2011), Skillful predictions of decadal trends in global mean surface temperature, *Geophys. Res. Lett.*, *38*, L22801, doi:10.1029/ 2011GL049508.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*, 84–88, doi:10.1038/nature06921.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997), A Pacific decadal climate oscillation with impacts on salmon, *Bull. Am. Meteorol. Soc.*, *78*, 1069–1079, doi:10.1175/1520-0477 (1997)078<1069:APICOW>2.0.CO;2.

Meehl, G. A., et al. (2009), Decadal prediction, can it be skillful?, *Bull. Am. Meteorol. Soc.*, *90*, 1467–1485, doi:10.1175/2009BAMS2778.1.

Meehl, G. A., A. Hu, and C. Tebaldi (2010), Decadal prediction in the Pacific region, *J. Clim.*, *23*, 2959–2973, doi:10.1175/2010JCLI3296.1.

Mochizuki, T., et al. (2010), Pacific decadal oscillation hindcasts relevant to near-term climate prediction, *Proc. Natl. Acad. Sci. U. S. A.*, *107*, 1833–1837, doi:10.1073/pnas.0906531107.

Mochizuki, T., T. Chikamoto, M. Kimoto, M. Ishii, H. Tatebe, Y. Komuro, T. Sakamoto, M. Watanabe, and M. Mori (2012), Decadal prediction

using a recent series of MIROC global climate models, *J. Meteorol. Soc. Jpn.*, *90A*, 373–383.

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke (2009), Initializing decadal climate predictions with the GECCO oceanic synthesis: Effect on the North Atlantic, *J. Clim.*, *22*, 3926–3938, doi:10.1175/2009JCLI2535.1.

Schlesinger, M. E., and N. Ramankutty (1994), An oscillation in the global climate system of period 65–70 years, *Nature*, *367*, 723–726, doi:10.1038/367723a0.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, *317*, 796–799, doi:10.1126/science.1139540.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*, 2283–2296, doi:10.1175/2007JCLI2100.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012, doi:10.1256/qj.04.176.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters and W. Hazeleger (2012), Skill in the trend and internal variability in a multi-model decadal prediction ensemble, *Clim. Dyn.*, in press.