Improvement of Initialized Decadal Predictions over the North Pacific Ocean by Systematic Anomaly Pattern Correction

HYE-MI KIM
School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York

YOO-GEUN HAM
Earth Systems and Environmental Sciences, Chonnam National University, Gwangju, South Korea

ADAM A. SCAIFE
Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received 21 August 2013, in final form 28 March 2014)

ABSTRACT

The prediction skill and errors in surface temperature anomalies in initialized decadal hindcasts from phase 5 of the Coupled Model Intercomparison Project (CMIP5) are assessed using six ocean–atmosphere coupled models initialized every year from 1961 to 2008. The initialized hindcasts show relatively high prediction skill over the regions where external forcing dominates, indicating that a large portion of the prediction skill is due to the long-term trend. After removing the linear trend, high prediction skill is shown near the centers of action of the dominant decadal climate oscillations, such as the Pacific decadal oscillation (PDO) and Atlantic multidecadal oscillation (AMO). Low prediction skill appears over the tropical and eastern North Pacific Ocean where the predicted anomaly patterns associated with the PDO are systematically different in model and observations. By statistically correcting those systematic errors using a stepwise pattern projection method (SPPM) based on the data in an independent training period, the prediction skill of sea surface temperature (SST) is greatly enhanced over the North Pacific Ocean. The SST prediction skill over the North Pacific Ocean after the SPPM error correction is as high as that over the North Atlantic Ocean. In addition, the prediction skill in a single model after correction exceeds the skill of the multimodel ensemble (MME) mean before correction, implying that the MME method is not as effective in addressing systematic errors as the SPPM correction.

1. Introduction

Prediction of decadal climate variability has become one of the major challenging issues in the climate science community because of its large socioeconomic impacts on the globe. While climate change projection focuses on reproducing the long-term trend, decadal prediction aims to predict the internally generated low-frequency variability superimposed on the externally forced climate change while perhaps also correcting forced signals through initialization (Smith et al. 2007; Meehl et al. 2009; Smith et al. 2012). To address the current status of decadal climate prediction in climate models, coordinated near-term climate prediction (or decadal prediction) experiments have been carried out under phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012).

Improvement of decadal prediction through initialization has been most clearly found over the North Atlantic Ocean, resulting in extended prediction skill out to the multiyear horizon due to decadal fluctuations of the Atlantic meridional overturning circulation (AMOC) (Keenlyside et al. 2008; Smith et al. 2010; van Oldenborgh et al. 2012; Doblas-Reyes et al. 2013; Goddard et al. 2013; Lienert and Doblas-Reyes 2013; Meehl et al. 2014; Yang et al. 2013; Ham et al. 2014). Over the North Pacific, some signs of improved prediction skill through initialization have been found associated with the Pacific decadal oscillation (PDO; Mantua
et al. 1997) or interdecadal Pacific oscillation (IPO; Keenlyside et al. 2008; Smith et al. 2010; Meehl et al. 2010; van Oldenborgh et al. 2012; Meehl and Teng 2012). Mochizuki et al. (2012) and Chikamoto et al. (2012) demonstrated that the models’ ability to follow the subsurface temperature evolution in the North Pacific increases through initialization. However, decadal prediction of the North Pacific appears to be far less skillful than prediction of the North Atlantic and there is much uncertainty around decadal variability mechanisms in the Pacific Ocean (Chikamoto et al. 2012; Guemas et al. 2012; Kim et al. 2012; Doblas-Reyes et al. 2013; Lienert and Doblas-Reyes 2013). Therefore, decadal prediction of the North Pacific climate variability remains a challenge.

Decadal prediction using climate models can be improved by realistic representation of processes and feedbacks in models following better understanding of the physical processes related to decadal variability, and by optimal assimilation methods for high-quality initial conditions. However, forecast error is still inevitable as small errors in the climate model or initial conditions grow and eventually destroy forecast skill. Therefore, efforts have been made to reduce model errors through postprocessing of decadal hindcasts. A multimodel ensemble (MME) is a useful approach to reduce model uncertainty and to enhance decadal prediction skill under limited computational resources (Chikamoto et al. 2012; Kim et al. 2012; Smith et al. 2012; Doblas-Reyes et al. 2013). However, this method is not effective when most climate models suffer from common systematic errors. This raises the need for further correction of anomalous systematic errors for decadal prediction, which has been successfully applied in seasonal prediction (Kang et al. 2004; Kang and Shukla 2006; Kim et al. 2008; Kug et al. 2008).

In this study, we attempt to correct the systematic errors in decadal hindcast anomalies, especially over the North Pacific Ocean where prediction skill is relatively low. Many previous studies used multimodel decadal hindcasts produced following the core experimental set (Taylor et al. 2012), where predictions start every five years, resulting in only 10 predictions during the period from 1961 to 2006. In this study, we assess hindcasts from additional decadal hindcast experiments initialized every year from 1961 to 2008, giving a total of 48 start dates.

Section 2 introduces details of hindcasts and observational data. Section 3 describes the decadal prediction skill and systematic errors in predicted anomalies. The skills of predictions before and after the statistical error correction are examined in section 4. A summary and discussion are provided in section 5.

2. Decadal hindcasts and observational data

We assess the CMIP5 decadal hindcasts conducted by six modeling centers: The UK Met Office decadal prediction system based on the Hadley Centre coupled global climate model, version 3 (HadCM3), Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1 (GFDL CM2.1), Model for Interdisciplinary Research on Climate, version 5 (MIROC5), Fourth Generation Canadian Coupled Global Climate Model (CanCM4), Goddard Earth Observing System version 5 (GEOS5), and Max Planck Institute Earth System Model, low resolution (MPI-ESM-LR, hereafter MPI-LR). Details of each model’s experimental configuration can be found in Meehl et al. (2014). The HadCM3, GFDL CM2.1, and CanCM4 experiments are based on 10-member ensembles, while MIROC5 has six members and GEOS5 and MPI-LR have three members. The hindcasts consist of ensembles of 10-yr retrospective predictions initialized every year from 1960/61 to 2007/08. There are more than just six modeling centers that have performed decadal hindcasts with a yearly frequency of start dates, but only the hindcasts from those six modeling centers consisted of 48-yr sets at the time of writing. The initialization date varies from November to the following January depending on the forecast system. The first forecast year starts in the first January of the hindcasts in this study. For GEOS5 hindcasts, the predictions initialized from 1982 to 1984 are not included because of technical problems (Ham et al. 2014). The annual mean surface temperature hindcasts are compared with observations of the surface air temperature from the Hadley Centre Climate Research Unit temperature dataset, version 4 (HadCRUT4; Morice et al. 2012). All hindcast data are interpolated to the horizontal resolution of observations, 5° longitude and 5° latitude.

Decadal hindcasts made by various groups are typically bias-adjusted by removing a time-independent climatological drift away from the initial state. The bias adjustments are first done as mentioned below in the conventional sense prior to the data analysis (i.e., removal of trend and application of anomaly error correction). In both observation and hindcasts, the anomaly is calculated by subtracting climatological means for the entire period from 1961 to 2008 following the method from the International CLIVAR Project Office (ICPO 2011). The model forecast anomaly is calculated as

\[ Y'_{ij} = Y_{ij} - Y_{j}, \]

where \( Y \) is the ensemble-average prediction, \( Y' \) is the anomaly of the raw forecast with respect to the forecast average \( Y_{j} \), \( j \) is the starting year (\( n = 1, 2, \ldots, 48 \)), and \( \tau \) is the forecast lead time (\( \tau = 1, 2, \ldots, 10 \)). The equally weighted average from ensemble members in each hindcast provides the values for
the ensemble mean. Temporal smoothing of hindcasts follows Goddard et al. (2013) for forecast lead time at 2–5- and 6–9-yr averages. The period of verification differs as the averaged lead time changes. The forecasts verified are those whose initial year is 1961 to 2008 \((n = 48)\) for 2–5-yr averaged prediction, and 1961–2004 \((n = 44)\) for the 6–9-yr average. A 4-yr running average is applied to the observed anomalies. The linear trend is calculated at each forecast lead time and then subtracted from the anomalies. We compare the prediction skill of the initialized hindcasts with persistence prediction, a commonly used statistical forecast, which assumes that future conditions will be the same as past conditions. For example, to calculate the prediction skill of SST anomalies for 2–5-yr averaged lead time, detrended SST anomalies of 2–5-yr averaged lead time are compared with the detrended SST anomaly of the initial year (year 0) over the entire period.

3. Prediction skill and systematic errors

a. Impact of the linear trend on decadal prediction skill

Relatively high prediction skill out to a decade and beyond in surface temperature appears over regions where external forcing dominates, indicating that a large portion of the decadal prediction skill is due to the long-term trend. (Keenlyside et al. 2008; Meehl et al. 2009; Pohlmann et al. 2009; Smith et al. 2010; Chikamoto et al. 2012; Kim et al. 2012; Mochizuki et al. 2012; van Oldenborgh et al. 2012; Goddard et al. 2013; Lienert and Doblas-Reyes 2013). In decadal predictions, the trend of global mean surface temperature tends to be stronger than the observed trend and this discrepancy increases as forecast lead time increases, resulting in a warm bias in decadal predictions (Smith et al. 2007; Kim et al. 2012; van Oldenborgh et al. 2012). Figure 1 shows the direct comparison of the modeled (MME) and observed linear trend. All MME results in the present analysis are based on the three best performing models that have relatively high prediction skill over the North Pacific Ocean (HadCM3, GFDL CM2.1, and MIROC5). The trend is calculated for the 4-yr moving average temperature anomaly in the observation and for the 2–5-yr averaged predicted temperature anomaly from the MME. The observed temperatures in the central and subtropical North Pacific do not show strong linear trends, while the predicted warming trend is apparent over the entire globe. The spatial pattern of the trend

![Figure 1](attachment:image_url)
could be slightly different in different observational datasets.

To examine additional prediction skill over the global warming signal, we calculate the residual of the linear trend and investigate the variability around the trend. Both in the observation and hindcasts, the long-term linear trend at each grid point is removed to avoid it dominating the prediction skill estimates. The detrended 2–5- and 6–9-yr averaged surface temperature anomaly is compared with the observed detrended 4-yr moving average surface temperature anomaly. Prediction skill is measured in terms of anomaly correlation coefficients and root-mean-square error (RMSE). The correlation coefficients and RMSE are calculated based on the ensemble mean for the hindcasts of each model. The pattern of the correlation coefficient of total skill with the linear trend included (Fig. 1c) is similar to the pattern of the linear trend in the observation (Fig. 1a), indicating that a large part of the prediction skill results from the linear trend. The prediction skill without trend (Fig. 1d) shows differences from the total skill. The highly predictable areas are shown near the centers of action associated with decadal climate variability in the major ocean basins. To examine the prediction skill of decadal hindcasts, we will focus on the surface temperature anomalies without the trend from now on.

Figure 2 shows the prediction skill (correlation coefficient) for 2–5-yr averaged surface temperature anomalies without the linear trend in the MME and in ensemble mean of each hindcast. In most of the models, high prediction skill appears over the central western North Pacific Ocean, tropical and subpolar North Atlantic, Indian Ocean, and Mediterranean regions, while low skill is seen in the eastern equatorial Pacific and North Pacific Ocean. The persistence prediction (Fig. 2h) shows positive correlation coefficients over most of the North Pacific. Figure 3 shows the RMSE for detrended surface temperature anomaly in the MME, ensemble mean of each hindcast, and persistence prediction. The RMSE is generally large where the correlation coefficient is low. The North Atlantic Ocean has smaller RMSE than that of the North Pacific Ocean. The RMSE is large over the high latitudes, the tropical Pacific, and the eastern and northern parts of the North Pacific in all hindcasts. The RMSE in persistence is generally larger than that in the hindcasts.

Generally, HadCM3, GFDL CM2.1, and MIROC5 show relatively higher skill than CanCM4, GEOS5, and MPI-LR in most of the Northern Hemisphere oceans, especially over the North Pacific. The relatively low prediction skill in CanCM4, GEOS5, and MPI-LR over the North Pacific could possibly be due to initialization problems or model biases (Müller et al. 2012; Ham et al. 2014), but the reason for lower prediction skill needs to be investigated in detail. Because of the low prediction skill over the Pacific Ocean, we exclude CanCM4, GEOS5, and MPI-LR in calculating the MME. Only predictions from HadCM3, GFDL CM2.1, and MIROC5 are used for calculating the MME with equal weighting on each ensemble member. The MME using three models (HadCM3, GFDL CM2.1, and MIROC5) exhibits better forecast skill than the MME using all six models (not shown), implying that the forecast skill is not just a function of the number of ensemble members. The difference of prediction skill between the three best models (HadCM3, GFDL CM2.1, and MIROC5) and three worst models (CanCM4, GEOS5, and MPI-LR) is significant with 95% confidence. Thus, a total of 26 ensemble members from the three best models are used for the MME.

b. Systematic errors over the North Pacific Ocean

We found that the prediction skill is relatively low in the eastern North Pacific and tropical Pacific in the three best models (HadCM3, GFDL CM2.1, and MIROC5). We hypothesize that the low skill is closely linked to the systematic error of representing the dominant mode of North Pacific climate variability. To examine errors of the predicted anomalies in the North Pacific Ocean, we investigate the dominant mode of sea surface temperature (SST) variability in both observation and predictions by performing empirical orthogonal function (EOF) analysis on detrended SST anomalies. The EOF analysis is performed on the 4-yr moving averaged detrended SST anomaly in observations over the North Pacific Ocean (10°–60°N, 110°–260°E). The first EOF explains 45.7% of the total detrended SST variability. The normalized principal component (PC) time series of the first EOF represents the PDO index (Fig. 4i). Figure 4a shows the regression of observed detrended SST onto the leading PC time series of observed SST anomalies (Fig. 4i). It is a “PDO-like” pattern in the Pacific, with negative anomalies in the interior North Pacific and positive anomalies along the eastern Pacific coast and tropical and subtropical Pacific (Fig. 4a).

The predicted pattern that is related to the observed PDO mode is obtained by the regression of the predicted detrended SST anomaly onto the normalized leading PC (Fig. 4i). The regressed SST patterns of the MME at 2–5-yr lead time (Fig. 4b) shows basinwide cooling over the North Pacific, whose pattern is similar to the predicted linear trend (Fig. 1b). The predicted pattern related to the observed PDO captures negative SST anomalies over the central North Pacific Ocean where the predictive skill is relatively high (Fig. 2a). However, the MME prediction shows negative anomalies...
over the eastern North Pacific and tropical Pacific, while the observed pattern exhibits positive anomalies. This is also true when the regression is performed on the individual models, especially for HadCM3, GFDL CM2.1, and MIROC5 (Figs. 4c–e).

Newman (2013) showed that the most notable deficiency in prediction of North Pacific decadal variability in initialized hindcasts appears to be related to the second eigenmode over the Pacific that represents decadal variability with sufficiently long e-folding time (Newman 2007), while the first mode corresponds to the trend, which has been removed in the present analysis. In the Pacific, the predicted SST patterns (Figs. 4b–e) resemble the second eigenmode of Newman (2007). This implies it is possible that the excessive excitation of the second mode in the forecast leads differences
between the observed and predicted first EOF. That is, the error in the dominant EOF of the forecast comes from the different contribution of predictable modes to the PDO rather than the model error in representing the PDO-related variability. This is supported by the notion that the PDO is the sum of independent red noise sources (Beran 1994; Newman 2007) rather than a single phenomenon. The different contribution of PDO-related modes is possibly due to differences in atmospheric variability (Hasselmann 1976), or air–sea coupling strength (Barsugli and Battisti 1998) in the simulation; however, further investigation should be performed to figure out the detailed mechanism.

Three models (HadCM3, GFDL CM2.1, and MIROC5) have common systematic errors in predicting the dominant mode of SST variability in the North Pacific. The regressed patterns from models (Fig. 4) are similar to the leading modes of the singular value decomposition (SVD) analysis between the observed and predicted SST (not shown) for HadCM3, GFDL CM2.1, and

![Fig. 3. As in Fig. 2, but for RMSE.](image-url)
Fig. 4. The regression of the SST anomaly (K) on the leading PC time series of observed SST anomalies: (a) observations, (b) MME, and (c)–(h) ensemble mean of each model at 2–5 yr average lead time. (i) The normalized PC time series from the first EOF of the observed SST whose value for a specific year indicates the averaged value for the following 4 yr. For example, the value in 1961 indicates the average of 1962–65.
MIROC5. This indicates that the three models share common errors in their simulation of decadal variability in the North Pacific Ocean. In particular, it is the general absence of oppositely signed anomalies in the eastern North Pacific Ocean. However, the leading modes of the other three models (CanCM4, GEOS5, and MPI-LR) are not strongly related to PDO like variability and the errors are not systematically different from observations (not shown), and thus the statistical error correction method cannot enhance their prediction skill. Therefore, we apply the statistical correction method to the predictions only from HadCM3, GFDL CM2.1, and MIROC5 whose predictions are systematically different from the observation.

4. Improvement of decadal predictions with statistical error correction

Systematic errors are seen over the North Pacific Ocean in the predicted detrended SST anomalies. Here, we apply the stepwise pattern projection method (SPPM; Kug et al. 2008) to correct the systematic error in the predicted anomalies.

a. Bias correction procedure

The SPPM is a pointwise regression model. Here we use the detrended SST anomaly at each grid point as both predictand and predictor. It has to be emphasized that the training period and validation period are distinct. The procedure for SPPM follows Kug et al. (2008) as

\[
\text{COV}(x) = \frac{1}{K} \sum_{t} O(t)P(x,t),
\]

\[
T(t) = \sum_{x} \text{COV}(x)P(x,t),
\]

\[
\alpha = \frac{1}{K} \sum_{t} O(t)T(t) - \frac{1}{K} \sum_{t} T(t)^2,
\]

\[
T(t_f) = \sum_{x} \text{COV}(x)P(x,t_f),
\]

\[
O(t_f) = \alpha T(t_f).
\]

Suppose that the predictand is \(O(t)\) and the predictor field is \(P(x,t)\). In the training period, \(O(t)\) (predictand) is an observed SST anomaly at a grid point and \(P(x,t)\) (predictor) is the predicted SST anomaly field from the hindcasts, and \(x\) and \(t\) are the spatial and temporal grid points. Over the training period \(K\), first we calculate the covariance pattern \(\text{COV}(x)\) between the predictand \(O(t)\) and predictor field \(P(x,t)\) in a certain domain \(D\). Then, the predictor field \(P(x,t)\) is projected onto the covariance pattern \(\text{COV}(x)\) to obtain a single time series \(T(t)\). Third, the regression coefficient \(\alpha\) is obtained by the time series \(T(t)\) and predictand \(O(t)\) over the training period \(K\). From these processes, we obtain a regression coefficient \(\alpha\) from the training period that will be applied to the forecast period later. The next processes are forecast procedures. Similar to how we obtain \(T(t)\), the time series \(T(t_f)\) can be obtained by projecting the predictor field \(P(x,t_f)\) in the forecast period onto the covariance pattern \(\text{COV}(x)\), which is already obtained from the training period. Finally, by multiplying the regression coefficient \(\alpha\) by \(T(t_f)\), the forecast SST anomaly \(O(t_f)\) at one grid point can be obtained. For verification of the predictions, we will compare the forecast SST anomaly \(O(t_f)\) with the observed SST anomaly in that grid point in the forecast period.

The selection of the predictor domain \(D\) is important. In general, the pattern projection models use a fixed domain. However, in the method suggested by Kug et al. (2008), the predictor domain is selected optimally based on the correlation between predictand and the predictor during the training period. To select the optimal predictor domain, we follow the same process as Kug et al. (2008). First, over the training period, the correlation coefficients between the \(O(t)\) and \(P(x,t)\) are calculated to search for the optimal domain among all possible grid points within a certain area (here, \(0^\circ-60^\circ\)N, \(135^\circ-250^\circ\)E). Grid points that have higher correlations than 0.3 are classified into several groups. The absolute correlation values are used as the criteria for grouping, ranging from 1 to 0.3 in 0.1 intervals. Initially, the grid points that exceed 0.9 are selected. If the number of grid points is less than 100, the grid points larger than 0.8 are added, and so on. The limit on the number of grid point (here 100) is arbitrary, but the results are not sensitive to the choice of correlation criterion or the minimum numbers of grid points, consistent with Kug et al. (2008).

The correction of the SST anomaly in MME prediction at grid point \(20^\circ\)N, \(120^\circ\)W (point A in Fig. 5) is performed as an example. This is a grid point where the models have apparent systematic errors (Fig. 4). In the training period, the observed SST anomaly at point A is \(O(t)\). Therefore, correlation coefficients between \(O(t)\) and predicted SST anomalies for 2–5 yr can be calculated (Fig. 5a). The correlation coefficients are obtained during the training period using predictions initialized from 1961 to 1985. Figure 5a shows a high negative correlation coefficient over the eastern and central North Pacific. It implies that the observed SST anomaly \(O(t)\) over point A is negatively correlated with the model-predicted SST anomaly. It is consistent with the results from the regressed SST pattern (Fig. 4) where
the dominant mode of the observed SST anomaly at point A shows opposite sign from the predicted SST anomaly over the North Pacific Ocean. The SST anomaly prediction at the point A can be improved by using those high-correlation areas as a predictor $P(x, t)$ rather than using the predicted anomaly directly. For better understanding, we calculate the correlation coefficient between observed SST anomaly $O(t)$ over point A with observed SST anomaly over the other grid points (Fig. 5b). The observed SST anomalies are positively correlated around the reference point A and resemble the PDO (Fig. 5b).

Based on the correlation coefficients, grid points (shaded area in Fig. 5a) are selected as predictors $P(x, t)$ to calculate the covariance pattern $\text{COV}(x)$. This process is applied to every grid point in the area $0^\circ–60^\circ$N, $135^\circ–250^\circ$E.

We use initial years 1961–85 (1961–80) as a training period and 1990–2008 (1990–2004) as a forecasting period for a 2–5-yr (6–9 yr) prediction. The SPPM is applied to the ensemble mean of each of the three models (HadCM3, GFDL CM2.1, and MIROC5) separately, and then averaged with equal weighting for the MME.

b. SST anomaly prediction with statistical error correction

The observed detrended SST anomaly (Fig. 6a) averaged over the period 1991–94, when the climate state stayed in the positive PDO phase, is shown together with predicted anomalies from 2–5- and 6–9-yr MME targeted to 1991–94 before correction (Figs. 6b,d). The SST anomaly patterns (Figs. 6b,d) resemble the regressed
pattern (Fig. 4b) with the minimum temperature in the North Pacific shifted slightly to the north compared to observations, and with a negative anomaly over the entire North Pacific Ocean. The observed positive anomaly in the eastern North Pacific and tropical central Pacific is not predicted (Figs. 6b,d). After the error correction with SPPM, the prediction has been greatly improved (Figs. 6c,e). Although the strong negative anomaly in the central North Pacific still has its maximum slightly shifted to the north, the positive SST anomaly in the eastern North Pacific surrounding the negative anomalies in the central North Pacific is captured, thus enhancing the prediction skill in those regions.

Figure 7 shows the prediction skill for the detrended SST anomaly before and after correction for 2–5-yr averaged lead time hindcasts from the MME and three individual models. Nineteen years (initial years of 1990–2008) from the forecasting period are used to calculate the temporal correlation coefficient. Before statistical correction, the MME shows low prediction skill in the eastern North Pacific and tropical central Pacific. After statistical correction, the areas of high prediction skill remain high or slightly improved (e.g., western Pacific), while the skill is enhanced over the central and eastern North Pacific (Fig. 7a). Note that the enhanced predictive skill in the North Pacific and tropical East Pacific after correction results from correcting systematic errors. The improvement is also clear in individual models (Figs. 7b–d), implying that the statistical correction has a positive impact as far as the predicted anomalies have systematic errors. Here, the SPPM has been applied to the ensemble mean of individual models (Figs. 7b–d) and then averaged over the...
predicted anomalies after correction for the MME (Fig. 7a). However, if the SPPM is applied to individual ensemble members rather than to the ensemble mean, the prediction skill does not improve significantly after correction.

Several sensitivity tests have been performed to test the robustness of the application of SPPM. We applied SPPM for forecasts with initial years 1990–2008 with training from forecast initial years 1961–80 and compared it with training from 1961 to 1985. We also applied the SPPM to the first half of the hindcast period based on training from the second half. The PDO-related anomalies are better predicted after the SPPM correction in all cases.

Figure 8 compares the area averaged detrended SST anomalies for 2–5 yr lead time over the eastern North Pacific (10°–30°N, 215°–250°E) and the central North Pacific (15°–40°N, 160°–210°E) before and after correction in the MME and GFDL CM2.1. We show the results from GFDL CM2.1 together with the MME as the

![Figure 7: Prediction skill (correlation coefficients) for the detrended SST anomaly between observations and model predictions at 2–5-yr average lead time (left) before and (right) after correction: (top to bottom) MME, HadCM3, GFDL CM2.1, and MIROC5.]
The prediction skill is improved the most after correction in these cases. Before correction, the ensemble mean SST anomaly over the eastern North Pacific predicted by the MME and GFDL CM2.1 does not capture the large positive anomalies in the early 1990s or the low-frequency variation with decreasing trend (Figs. 8a,b). After the error correction, the anomalies resemble the observations, with positive anomalies in the 1990s, a decreasing trend up to 2000, and a slight increase between 2001 and 2005 (Figs. 8a,b). In the MME, the temporal correlation coefficient between the predicted and observed SST anomaly over the forecast period is $-0.70$ before correction and $0.76$ after error correction. The predictions over the central North Pacific also show improvements after correction (Figs. 8c,d).

The probability of the spatial frequency distributions of pointwise correlation coefficients before and after correction is shown in Fig. 9. The statistics are based on all grid points over the entire inner domain over the North Pacific ($15^\circ$–$45^\circ$N, $120^\circ$–$245^\circ$E) with a total of 26 ensemble members from three models (HadCM3, GFDL CM2.1, and MIROC5). The frequency is calculated with 0.1 intervals and then divided by the total numbers of grid points over the selected domain to represent the probability of occurrence. Before correction, correlation coefficients are almost normally distributed with symmetry about zero mean (Fig. 9, black solid line). The peak frequency is shifted to high correlation after the correction (Fig. 9, red bar), indicating that the prediction has higher skill over the North Pacific than before correction. After correction, the peak has shifted from zero to 0.2 and the probability of correlations above 0.5 has almost doubled.

Statistical significance for differences in distributions is determined using a two-sample Kolmogorov–Smirnov test (Smirnov 1948). To calculate the independent number of samples, the effective number of grid point is measured using an autocorrelation map. We chose a single point (i.e., $30^\circ$N, $185^\circ$E) where the skill has mostly improved, then calculated the autocorrelation map between the chosen temperature anomaly and anomalies of the whole grid within the target region to measure the effective radius. The effective radius is defined as the length where the autocorrelation is at 99% confidence level (i.e., correlation value 0.57), and we assume that two grid points separated by more than the effective radius are independent of each other. Finally, the total number of independent grid points in the target region ($15^\circ$–$45^\circ$N, $120^\circ$–$245^\circ$E) is defined by dividing the total area length by the effective radius. For example, if the effective radius in longitudinal direction is 10 degrees, the number of independent grid points in the longitudinal direction is 12.5 [i.e., $(245 - 120)/10 = 12.5$]. A two-sample Kolmogorov–Smirnov test shows that the distributions before and after correction are significantly different from each other at the 90% confidence level. It could be argued that the 19-yr period used for validation of the statistical correction might be too short. To address this issue, we change the length of the training and forecasting periods to 1961–75 and 1980–2008, respectively. The result is similar to Fig. 9.

This study shows that the statistical correction improves deterministic prediction skill of multiyear forecasts. However, in terms of probabilistic forecast, we found that the ensemble spread is underestimated compared to
the forecast error when the statistical correction method is applied. This is due to the individual ensemble member being corrected too much through the statistical correction, and the differences among ensemble members become too close to each other.

5. Summary and discussion

The prediction skill and errors for the predicted surface temperature anomalies in CMIP5 decadal hindcasts have been assessed. Six ocean–atmosphere coupled models that are initialized every year from 1961 to 2008 and the multimodel ensemble mean of the models are compared in parallel. The prediction skill is examined by detrending the predicted anomalies. The highly predictable areas are shown to be near the centers of action of the dominant decadal climate oscillations, the PDO and AMO, while lower prediction skill appears over the tropical and subtropical North Pacific and eastern North Pacific Ocean where three models (HadCM3, GFDL CM2.1, and MIROC5) have common systematic errors in the predicted SST anomaly patterns. These three models are capable of predicting a large-scale PDO signal, although the spatial patterns associated with the PDO are systematically different from the observed pattern. By statistically correcting the systematic pattern errors using the SPPM method, based on independent data for training and forecasting periods, the prediction skill is enhanced over the North Pacific Ocean.

This study implies that the statistical error correction method can be applied successfully to decadal predictions, even though decadal prediction tends to have a smaller number of samples for the training period than seasonal or weather prediction. This statistical error correction leads to better results than the MME method, which has been known as a useful approach to enhance decadal prediction skill under limited computational resources. To examine this point in more detail, the

![Figure 9](image1.png)

**Fig. 9.** Spatial frequency distribution of correlation coefficients before (line) and after (bars) error correction for a 2–5-yr average lead time. The frequency is calculated over the North Pacific Ocean (120°–245°E, 15°–45°N) for all ensemble members from HadCM3, GFDL CM2.1, and MIROC5. A two-sample Kolmogorov–Smirnov test shows that the distributions before and after error correction are significantly different from each other at the 90% confidence level.

![Figure 10](image2.png)

**Fig. 10.** The mean prediction skill (correlation coefficients) for 2–5 yr of SST predictions averaged over the North Pacific Ocean (120°–245°E, 15°–45°N) in the ensemble mean of each model and the MME for before (blue) and after (red) correction.
prediction skills for 2–5-yr averaged SST anomaly over the North Pacific Ocean (15°–45°N, 120°–245°E) are compared (Fig. 10). Before correction, the individual models show maximum correlations around 0.2. Although the three models are averaged with equal weighting, the skill for the MME is still around 0.2. However, after applying the statistical correction, each prediction shows enhancement of prediction skill. The correlation in single model after correction exceeds the skill of the MME before correction. It clearly shows that the MME method is not effective because it cannot remove common systematic errors, while the statistical correction has a substantial impact. On the other hand, the MME method is effective in reducing other forecast errors from the models and initial condition, which are not common to different models, while statistical correction is not. In addition, in the uninitialized simulations, the spatial structure and magnitude of PDO is generally realistic in the coupled climate models (Deser et al. 2012), while our results of initialized simulation show systematic errors of the PDO-related pattern. Therefore, to enhance the ability of multiyear prediction over the North Pacific Ocean, the dynamic origins of the errors in the initialized hindcasts and the amplification of the error by forecast lead time need to be further investigated.

It is shown that the statistical error correction is beneficial for decadal prediction over the North Pacific Ocean if models have common errors in simulating Pacific decadal variability. We found that the statistical error correction over the North Atlantic Ocean is not as effective as the North Pacific Ocean because most of the models already predict the Atlantic decadal climate variability relatively well and because there is no strong systematic error in predictions over North Atlantic SST. As a result, after the statistical correction, the prediction skill of SST over the North Pacific becomes similar to that over the North Atlantic. For example, the SPPM method greatly enhances the mean prediction skill over the North Pacific (correlation coefficients of 0.45 over 15°–45°N, 120°–245°E), which is comparable to that over the North Atlantic (correlation coefficients of 0.42 over 0°–60°N, 280°–350°E before the correction, and similar scores after the correction).

Given that there is currently a consensus that the prediction skill over the Pacific Ocean is less than that over the North Atlantic Ocean, and this may limit the usage and application of decadal forecasts to the Atlantic sector, this study provides a foundation to extend the utility of decadal forecasts to the global oceans including the Pacific Ocean and its surrounding countries over the Asia–Pacific sectors.

Acknowledgments. Kim was supported by the Korea Meteorological Administration Research and Development Program under Grant APCC 2013-3141 and by the New York Sea Grant R/MARR14NJ-NY. Adam Scaife was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). We thank Dr. J. S. Kug for providing us the SPPM code and Daehyun Kang (UNIST) for providing downloaded CMIP5 decadal hindcast data. The constructive and valuable comments of two anonymous reviewers are greatly appreciated.

REFERENCES


