An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC

Yoshimitsu Chikamoto · Masahide Kimoto · Masayoshi Ishii · Takashi Mochizuki ·
Takashi T. Sakamoto · Hiroaki Tatebe · Yoshiki Komuro · Masahiro Watanabe ·
Toru Nozawa · Hideo Shiogama · Masato Mori · Sayaka Yasunaka · Yukiko Imada

Abstract Decadal climate predictability is examined in hindcast experiments by a multi-model ensemble using three versions of the coupled atmosphere-ocean model MIROC. In these hindcast experiments, initial conditions are obtained from an anomaly assimilation procedure using the observed oceanic temperature and salinity with prescribed natural and anthropogenic forcings on the basis of the historical data and future emission scenarios in the Intergovernmental Panel of Climate Change. Results of the multi-model ensemble in our hindcast experiments show that predictability of surface air temperature (SAT) anomalies on decadal timescales mostly originates from externally forced variability. Although the predictable component of internally generated variability has considerably smaller SAT variance than that of externally forced variability, ocean subsurface temperature variability has predictive skills over almost a decade, particularly in the North Pacific and the North Atlantic where dominant signals associated with Pacific decadal oscillation (PDO) and the Atlantic multidecadal oscillation (AMO) are observed. Initialization enhances the predictive skills of AMO and PDO indices and slightly improves those of global mean temperature anomalies. Improvement of these predictive skills in the multi-model ensemble is higher than that in a single-model ensemble.

Keywords Decadal climate prediction · Multi-model ensemble · AMO · PDO

1 Introduction

Regional climate predictions for the coming decade have gained considerable attention since they can help to solve socioeconomic problems owing to climate change (Hibbard et al. 2007; Cox and Stephenson 2007). These predictions are one of the most challenging issues faced by the climate community (Murphy et al. 2010). To predict regional climate variability on decadal timescales, we need to consider the internally generated atmospheric and oceanic variabilities as well as externally forced variability related to global warming (Hawkins and Sutton 2009). Using a coupled atmosphere-ocean general circulation model (AOGCM) with initialization technique for assimilating observation in the ocean and prescribed natural and anthropogenic forcings (e.g., greenhouse gases and aerosol concentrations, solar cycle variations, and major volcanic eruptions), a recent series of studies has successfully predicted regional climate variability on decadal timescales (Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009; Mochizuki et al. 2010; Chikamoto et al. 2012). In order to provide useful future climate information for the

Y. Chikamoto (✉) · M. Kimoto · M. Watanabe · M. Mori
Atmosphere and Ocean Research Institute, The University
of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi,
Chiba 277-8568, Japan
e-mail: chika44@aori.u-tokyo.ac.jp

M. Ishii
Meteorological Research Institute, Tsukuba, Japan

M. Ishii · T. Mochizuki · T. T. Sakamoto · H. Tatebe ·
Y. Komuro
Japan Agency for Marine-Earth Science and Technology,
Yokohama, Japan

T. Nozawa · H. Shiogama · S. Yasunaka
National Institute for Environmental Studies, Tsukuba, Japan

Y. Imada
Department of Civil and Environmental Engineering, Tokyo
Institute of Technology, Tokyo, Japan

1202																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																Y. Chikamoto et al.

coming decades, further improvement of predictive skills on decadal timescales is an important topic in the climate community.

Experiments of decadal climate prediction are restricted to computational resources because such studies need to perform ensemble model integrations over long-term periods utilizing several types of experiments designed by the Coupled Model Intercomparison Project-5 (CMIP5) (Taylor et al. 2009; Murphy et al. 2010). When we increase the resolution of AOGCM, model performance may improve because of a more realistic representation of small-scale phenomena (Sakamoto et al. 2012; Delworth et al. 2011), which may have a positive impact on the improvement of predictive skills. Nevertheless, increase in the resolution using limited computational resources leads us to reduce the number of ensemble members which may underestimate the ensemble spread during prediction. In addition, for decadal climate prediction, definition of a model climate drift during prediction, which may disrupt the prediction, requires a large number of hindcast samples coupled with a large number of ensemble members. Decadal climate predictability is also affected by uncertainty due to model deficiencies such as incorrect representation of physics, climatological field, internal variability, externally forced variability, and climate sensitivity.

A multi-model ensemble performed by many research centers is a useful approach to reduce model uncertainty under limited computational resources. For example, Jin et al. (2008) showed that the El Niño-Southern Oscillation (ENSO) predictive skill of a multi-model ensemble with ten different AOGCMs is higher than those of any individual model and hindcast experiments based on persistence and a dynamic statistical model. van Oldenborgh et al. (2012) investigated the decadal predictive skill in hindcast experiments using a multi-model ensemble of four European AOGCMs and indicated that sea surface temperature (SST) in the North Atlantic and Pacific Oceans regionally has the predictive skills beyond the persistence. Among these hindcast experiments using four AOGCMs, one group adopted a strategy of anomaly initialization (i.e., the observed anomalies are incorporated into the model anomalies) in order to reduce the model climate drift during the experiments, whereas the other three groups initialized the predictions by using full-field observations in order to provide initial conditions most similar to the real climate system (Doblas-Reyes et al. 2010). These studies suggest that the multi-model ensemble approach is an effective tool for decadal climate prediction.

In this study, we examine predictable components on decadal timescales by using the multi-model ensemble approach and discuss the number of ensemble members suitable for decadal climate prediction. In addition, we address the impact of anomaly initialization for decadal

climate prediction by comparing the skills of initialized and uninitialized predictions. To reduce the model climate drift during prediction, we conducted decadal climate prediction experiments on the basis of anomaly initialization using AOGCMs with low, medium, and high resolutions (Tatebe et al. 2012; Mochizuki et al. 2012; Chikamoto et al. 2012).

The remainder of this paper is organized as follows. Section 2 describes the three versions of our model and experiments to assess the predictability of decadal climate variability. In Sect. 3, we evaluate the geographical distribution of the predictable area related to surface air and oceanic subsurface temperature anomalies. The predictability of the major climate variability on decadal timescales in terms of the global mean temperature, Atlantic multidecadal oscillation (AMO), and Pacific decadal oscillation (PDO) is examined in Sect. 4. We conclude and discuss the study in Sect. 5.

2 Model and experiments

2.1 Model

In this study, we use the AOGCM MIROC that was cooperatively developed by the Atmosphere and Ocean Research Institute of The University of Tokyo, National Institute for Environmental Studies (NIES), and Japan Agency for Marine-Earth Science and Technology. MIROC's atmospheric component is based on a global spectral dynamical core and a standard physics package (K-1 Model Developers 2004) developed at the Center for Climate System Research (CCSR)/NIES/Frontier Research Center for Global Change (FRCGC). The ocean general circulation model is the CCSR ocean component model (COCO; Hasumi 2006) which includes a sea ice model (K-1 Model Developers 2004; Komuro et al. 2012). No correction is applied in exchanging heat, water, and momentum fluxes between the atmosphere and the ocean. In addition, a land model that incorporates a river module is used (Takata et al. 2003; K-1 Model Developers 2004).

As summarized in Table 1, we used three different versions of MIROC: MIROC3m, MIROC4h, and MIROC5. In MIROC3m, the atmospheric and oceanic resolutions are a T42 spectral model with 20 levels on vertical σ-coordinates and an approximately $1° \times 1°$ longitude-latitude grid with 44 vertical levels, respectively. MIROC4h includes the same model physics as MIROC3m but at an eddy-permitting resolution for the ocean: a T213 spectral model with 56 levels in the atmosphere and an approximately $1/4° \times 1/6°$ longitude-latitude grid with 47 vertical levels in the ocean. In MIROC5, most parts of model physics are updated or replaced with new parameterization schemes derived from MIROC3m and

Table 1 Summary of MIROC versions

Version	Atm. res.	Ocn. res.	Reference
MIROC3m	T42L20	$1° \times 1°$ grid, L44	Nozawa et al. (2007)
MIROC4h	T213L56	$1/4° \times 1/6°$ grid, L47	Sakamoto et al. (2012)
MIROC5	T85L40	$1° \times 1°$ grid, L49	Watanabe et al. (2010)

MIROC4h. The resolution of MIROC5 is T85 spectral with 40 sigma-pressure hybrid vertical levels in the atmosphere and approximately 1° horizontal grid of curvilinear coordinates with 49 vertical levels in the ocean. Details of the performance and settings of MIROC3m, MIROC4h, and MIROC5 are described in Nozawa et al. (2007), Sakamoto et al. (2012), and Watanabe et al. (2010), respectively.

2.2 Experiments

In order to evaluate decadal climate predictability, we conducted the following three experiments using the three versions of MIROC: the twentieth century climate simulation (NoAS), the data assimilation (ASSM), and the hindcast (HCST) experiments. In the NoAS experiment, the model was prescribed by historical natural and anthropogenic forcings such as greenhouse gases and aerosol concentrations, solar cycle variations, major volcanic eruptions and future emission scenarios based on the report of Intergovernmental Panel of Climate Change (IPCC). Using the model climatology defined by the NoAS experiment, the observed temperature and salinity anomalies in the ocean (Ishii et al. 2006; Ishii and Kimoto 2009) were incorporated into model anomalies by an incremental analysis update (IAU) scheme (Bloom et al. 1996; Huang et al. 2002) in the ASSM experiment. The observed temperature and salinity anomalies, obtained from the gridded monthly objective analysis covering the whole ocean produced by Ishii et al. (2006) and Ishii and Kimoto (2009), were linearly interpolated to each day and to the ocean model grid. Analysis increment estimated from a temporally and spatially invariant model-to-observation ratio in analysis errors was incorporated into the model as constant forcing in the model tendency equation during an analysis interval of one day. Under our current computational resources, our simple assimilation approach can be applied for decadal climate prediction. From the ASSM experiment, we can first obtain a pair of atmospheric and oceanic initial conditions and then conduct ensemble HCST experiments. As described below and summarized in Tables 2 and 3, the experimental designs and the number of ensemble members are different among experiments in MIROC3m, MIROC4h, and MIROC5. Details of the model experiment and the assimilation procedure are described in Tatebe et al. (2012) and Mochizuki et al. (2012).

In MIROC3m, the NoAS, ASSM, and HCST experiments are based on 10-member ensembles (Table 3). Nozawa et al. (2005) and Shiogama et al. (2007) performed 10-member ensemble simulations for the 1850–2000 period using the historical data of natural and anthropogenic forcings and the 2001–2030 period using the A1B-type emissions scenario of the Special Report on Emissions Scenarios (Nakicenovic et al. 2000). These ensemble simulations are started from different initial states obtained from the pre-industrial control run and correspond to the NoAS experiment in MIROC3m. Therefore, each member of the NoAS experiment has a different phase of internally generated variability such as ENSO, PDO, and AMO. From each member of the NoAS experiment, we obtained 10-member initial states on January 1, 1945, and conducted 10-member ensemble assimilations over the 1945–2009 period. During data assimilation, the observed temperature and salinity anomalies in the upper 700-m ocean depth (Ishii et al. 2006) are incorporated into MIROC. Although each ensemble member in the ASSM experiment tends to have a same phase of oceanic internally generated variability, our simple assimilation approach would have different trajectories among ensemble members, particularly in the atmospheric high-frequency states. On the basis of the 10-member assimilation experiment, we also obtained 10-member initial states and conducted 10 sets of 11-year-long, 10-member ensemble hindcast experiments initialized on January 1, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, and 2005. These experimental designs in MIROC3m are the same as those in Mochizuki et al. (2010).

In MIROC5, 3-member ensemble simulations in the NoAS experiment were conducted for the 1850–2005 period using the historical forcing dataset updated for CMIP5 and for the 2006–2100 period using the emissions scenario of the Representative Concentration Pathway (RCP4.5; Moss et al. 2008, 2010). Using two initial conditions on January 1, 1945, taken from 3-member simulations of the NoAS experiment, we conducted 2-member assimilation experiment for the 1945–2007 period. In the ASSM experiment, the observed temperature and salinity anomalies from the surface to 3,000-m depth (Ishii and Kimoto 2009) were assimilated into MIROC5 in a similar manner to MIROC3m. In the HCST experiment, we conducted 10 sets of 10-year, six-member ensemble hindcast experiments initialized in 1961, 1966, 1971, 1976, 1981,

Table 2 Summary of our experimental design for the three MIROC versions

	Ext. for.	Obs. depth	ASSM period	Initial
MIROC3m	20C3M & A1B	700-m	1945–2009	Every 5-year from 1960
MIROC4h	20C3M & RCP4.5	3,000-m	1955–2007	Every 5-year from 1961
MIROC5	20C3M & RCP4.5	3,000-m	1945–2007	Every 5-year from 1961

Table 3 Number of ensemble members in each experiment

	MIROC3m	MIROC4h	MIROC5
NoAS	10	3	3
ASSM	10	1	2
HCST	10	3 (LAF)	6 (LAF)

Ensemble members of hindcast experiments in MIROC4h and MIROC5 were obtained from the lagged average forecast (LAF) with previous 3- and 6-month lags: initial conditions were July 1, October 1, and January 1

1986, 1991, 1996, 2001, and 2006. Six ensemble members were obtained from the lagged average forecasting (LAF) method with the previous 3- and 6-month lags on the basis of the two-member ASSM experiment. For example, for the 1961 initialization, predictions commenced on July 1, 1960, October 1, 1960, and January 1, 1961.

The experimental design of MIROC4h is almost the same as that of MIROC5. The main differences between these experimental designs are the periods in the NoAS and ASSM experiments, the number of ensemble members, and the assimilation procedure. In MIROC4h, the NoAS experiment has three ensemble members for 1950–2007 but one member for 2008–2035. To reduce computational cost, the ASSM experiment was conducted with one member for 1950–2007. In terms of horizontal resolution, a small-scale eddy in the ocean is resolved in MIROC4h but not in the observation (Ishii and Kimoto 2009). When we assimilate the observation on a $1° \times 1°$ horizontal grid into the eddy resolving model such as MIROC4h, analysis increment estimated by our system adopted in MIROC3m and MIROC5 acts to reduce the small-scale eddy resolved by MIROC4h. To retain the small-scale eddy, large-scale components of analysis increment estimated from a spatially smoothed field is incorporated into MIROC4h (Tatebe et al. 2012).

To distinguish climate signals from unrealistic signals arising from the model climate drift during prediction, drifts estimated in each model were excluded in the HCST experiment on the basis of INTERNATIONAL CLIVAR PROJECT OFFICE (2011) and Chikamoto et al. (2012). In each model, differences between the model predicted anomaly at a forecast lead time and the corresponding observational anomaly at a valid time are averaged over 9 sets initialized from 1960 or 1961 to 2000 or 2001 every 5 years. As shown in Chikamoto et al. (2012), the HCST

experiment in MIROC5 displays a temperature drift during prediction, such as El Niño, for the initial 3-year lead time while a small temperature drift appears in the North Atlantic in MIROC3m. Although we used anomaly assimilation techniques to reduce the model climate drift, an inconsistency in the thermocline depth between the observation and the model may cause a model drift. Dynamical processes that induce the model climate drift during prediction will be reported elsewhere. The multi-model ensemble approach may reduce these systematic biases, as explained in Sect. 3.

After removing model drift during prediction in each model, multi-model ensemble in the HCST experiment is obtained from an average over three ensemble means in each model. Although initial dates starting from initial conditions depend on ensemble members and model versions as shown above, forecast lead time is defined as departure from initial dates on January 1, 1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, and 2006. In other words, the HCST experiment in MIROC3m are regarded as the LAF method at 1-year lag. The anomalies are defined as departures from the climatological mean value for 1971–2000 period in each model and are linearly interpolated into MIROC3m grid.

3 Overview of decadal climate predictability

To examine the regional skill of surface air temperature (SAT) anomalies in decadal climate prediction, we computed the centered anomaly correlation coefficient (ACC) and root-mean-squared error (RMSE) of anomalies between the National Centers for Environmental Prediction (NCEP) reanalysis (Kalnay et al. 1996) and the multi-model ensemble of the HCST experiments for 1-, 2–4-, and 5–9-year lead times (Fig. 1). Highly correlated and lower RMSE regions appeared mainly over the Indian Ocean, the West Pacific, and the North Atlantic. For the 1- and 2–4-year lead times, RMSEs in the HCST experiments were almost entirely lower than that in the NoAS experiments (hatched orange region in Fig. 1b, d). For the following 5–9-year lead time, however, the lower RMSE region in the HCST experiments diminished considerably. These lower RMSEs suggest that initialization contributes to the enhancement of the predictive skill of SAT anomalies for initial several years, while its impact becomes unclear later

in our current hindcast system. Despite the fact that the model climate still imperfectly represents the observed climate state, our results in a grid-to-grid comparison imply that some regional climate changes are predictable (i.e., ACC exceeds the statistical significance at the 90 % level) in our multi-model ensemble.

In order to obtain potentially predictable variations in the model and observational fields separately, we apply a singular value decomposition (SVD) analysis between the hindcast system and the observations. The SVD analysis (Wallace et al. 1992; Bretherton et al. 1992) can provide the observed and the model represented spatiotemporal structures correlated between two fields (i.e., maximized the cross-covariance between two fields). We expect that the lower-order modes of SVD analysis for these two fields could capture predictable components in our hindcast experiment through maximized cross-covariance, even though model simulated variations may have different structures from the observations due to model deficiencies. Figure 2 shows the first and second SVD modes of SAT anomalies between the multi-model ensemble of the HCST experiment for a 1–3-year lead time and the NCEP reanalysis for its valid time. Although we generally apply the SVD analysis for dynamically related variables such as SST and geopotential hight anomalies (Wallace et al. 1992), there are no dynamical connection between the observational and hindcasted fields. In our paper, we calculated homogeneous correlation maps in Fig. 2 in order to consider the observational and model fields separately.

In the first SVD mode (left panels in Fig. 2), SAT globally increased for the recent decades in the observation and the multi-model ensemble, which seems to represent a global warming signal associated with external forcing. The first SVD mode in the observed SAT anomalies shows increasing trends over the Western tropical Pacific, the Indian Ocean, the North Atlantic, and the Arctic Ocean but no or decreasing trends over some parts of the Eastern North Pacific, the Southern Ocean, and the Tibetan Plateau (Fig. 2a). These observed temperature trends are qualitatively consistent with the global warming signals estimated from linear trend (Fig. 3a) and regression maps with global mean SAT (Fig. 3c) over 1955–2010 period. In fact, uncentered pattern correlation coefficients of the first SVD mode in observation (Fig. 2a) with the linear trend (Fig. 3a) and the regression map (Fig. 3c) are 0.77 and 0.79, respectively. On the other hand, the first SVD mode in the HCST experiment has a zonally uniform structure of SAT increasing trend except for some regions of the Southern Ocean (Fig. 2c). When externally forced variations of SAT anomalies are estimated from linear trend and regression map with global mean SAT in the NoAS experiment (Fig. 3b, d), SAT increasing pattern has zonally uniform structures except for some parts of the Southern

Ocean. Moreover, those patterns of first SVD modes in the HCST experiment and observations are almost identical to those obtained by applying SVD analysis between the NoAS experiment and observations (Fig. 3e, f): uncentered pattern correlation coefficients of these first SVD modes in observations (Figs. 2a, 3e) and models (Figs. 2c, 3f) are 0.99 and 0.98, respectively.

The second SVD mode shows a mixture of PDO-like pattern in the Pacific and the AMO-like pattern in the North Atlantic (right panels in Fig. 2). In the Pacific region, observed SAT anomalies become positive in mid-latitude regions of the North and South Pacific and negative in the tropical Pacific and the high-latitude region of southern Pacific. This SAT pattern in observation is similar to the regression map on the PDO index (definition of PDO index is described in Sect. 4.3): the uncentered pattern correlation coefficient is −0.74 over the Pacific region (50°S–70°N, 100°E–70°W). In the Atlantic, positive SAT anomalies in observation are limited around Greenland, which appears at 5 years later after the mature phase of AMO. In fact, uncentered pattern correlation coefficients between the second SVD mode and the regression map of SAT on the AMO index are 0.47 simultaneously but increases up to 0.63 at the 5-year lag over the Atlantic region (50°S–70°N, 60°W–0°). These observed signals in the Pacific and Atlantic regions also appear in the second SVD mode of our multi-model ensemble (Fig. 2d). As pointed out by Solomon et al. (2010), distinguishing between internally generated and externally forced variations is a difficult problem due to some assumption intrinsic to an analysis technique. Our SVD analysis is one of the useful tools to roughly estimate predictability associated with natural and externally forced variations in observational and model represented fields under the small sampling such as our HCST experiment (10 samples in this paper).

The first and second SVD modes of SAT anomalies explain 97 and 1.4 % of the total covariance between the observation and the multi-model ensemble. In other words, the most predictable component of SAT anomalies in our current hindcast system is the global warming signal. When we focus on the ocean subsurface temperature variability instead of SAT, the covariance ratio explained by the second SVD mode considerably increases (to 20 %) as shown by Mochizuki et al. (2012). This high covariance ratio suggests that ocean subsurface temperature variability and the global warming signal are also important for the long-term predictability of SAT anomalies.

Figure 4 shows predictive skill maps for vertically averaged ocean temperature anomalies from the surface to 300-m depth (VAT300) using objective analysis produced by Ishii and Kimoto (2009). In the North Pacific, significant predictability (i.e., high ACC and low RMSE) appears particularly around the Kuroshio–Oyashio extension and
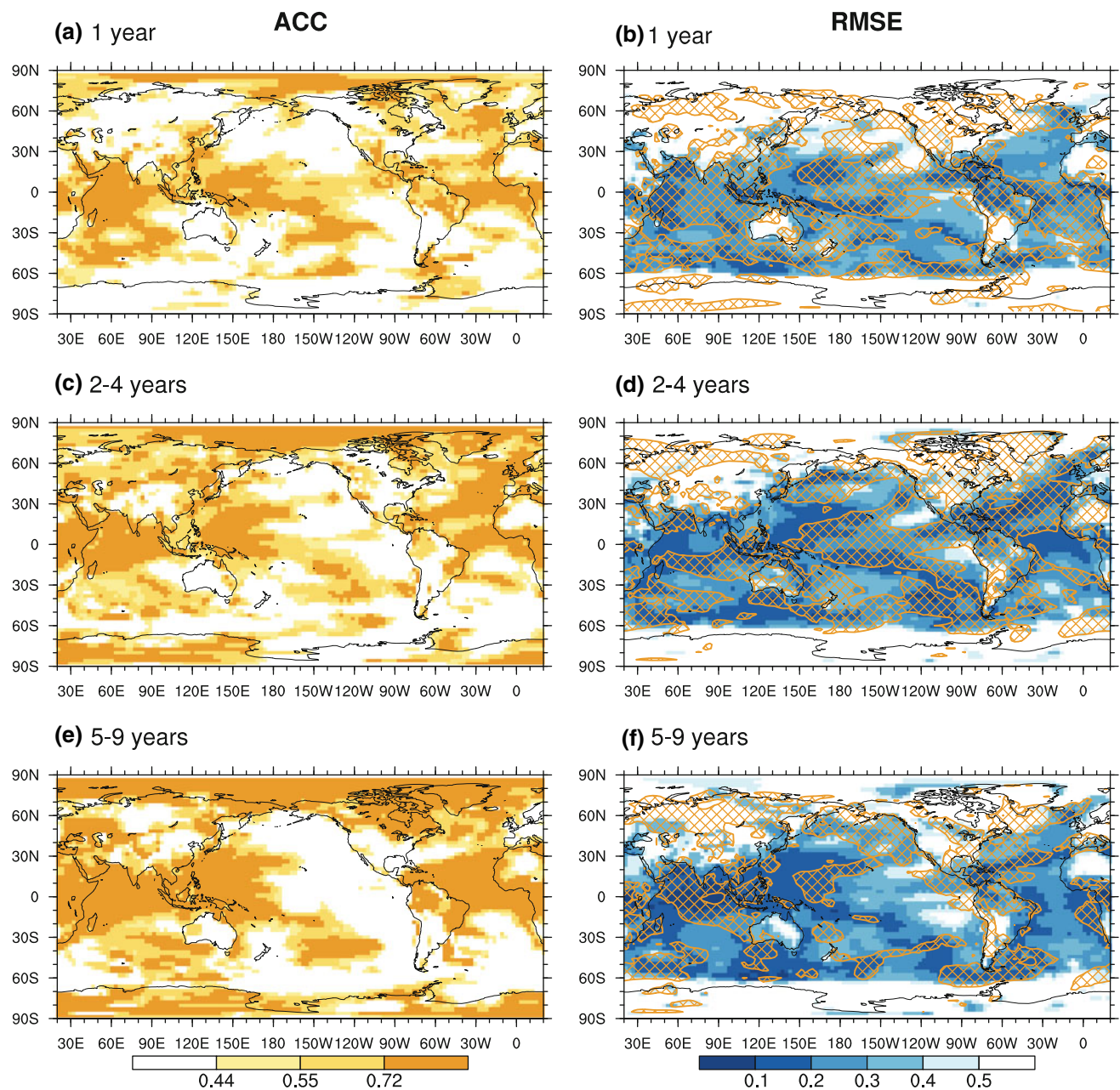
Fig. 1 Surface air temperature predictability of the multi-model ensemble for **a**, **b** 1-, **c**, **d** 2–4-, and **e**, **f** 5–9-year lead times. The predictive skill is measured according to the ACC (*left panels*) and RMSE (*right panels*) between the NCEP reanalysis (Kalnay et al. 1996) and multi-model ensemble of the HCST experiments. The multi-model ensemble is computed by averaging over three ensemble means in each model. Correlation coefficients of 0.44, 0.55, and 0.72

correspond to the statistical significance at the 90, 95, and 99 % levels, respectively, by using a one-sided Student *t* test with eight degrees of freedom. Hatching orange region in right panels indicates the regions where the ratio of RMSE in the HCST experiments to that in the NoAS experiments is less than 0.9, implying the skill improvement of 10 % by initialization compared to an uninitialized run

subtropical oceanic frontal regions (Fig. 4), which are the centers of action of the PDO pattern (Trenberth and Hurrell 1994; Mantua et al. 1997; Mochizuki et al. 2010), while the predictable component of SAT is limited in the western part of the North Pacific (Fig. 1). Moreover, the predictive skills of VAT300 are improved in the HCST experiment compared with those in the NoAS experiment in those

regions (hatched orange region in Fig. 4). In addition, ACC exceeding 0.72 in VAT300 extends more westward from the west coasts of Europe and North Africa than in SAT. In particular, lower RMSEs in the HCST experiment than those in the NoAS experiment extend to the Labrador and Greenland Seas, where the strength of deep convection can force meridional overturning circulation and then may be

SVD for global SAT

SVD1, SCF = 97.2%, R = 0.98   **SVD2, SCF = 1.3%, R = 0.95**
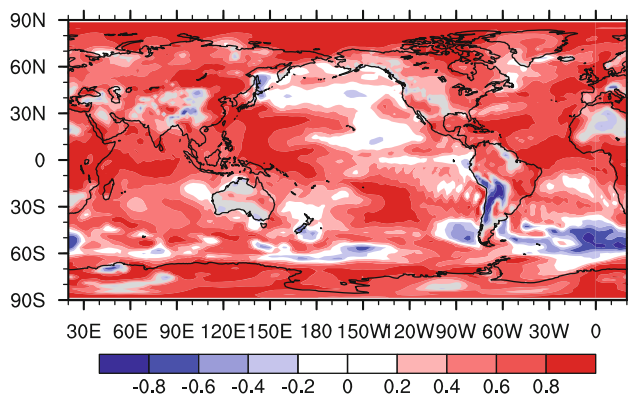
(a) OBS (NCEP)   **(b)** OBS (NCEP)
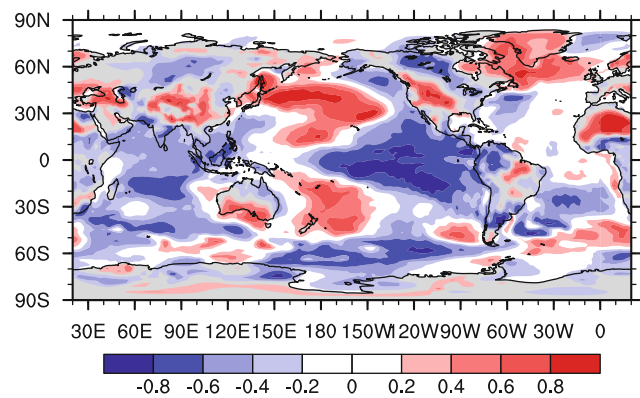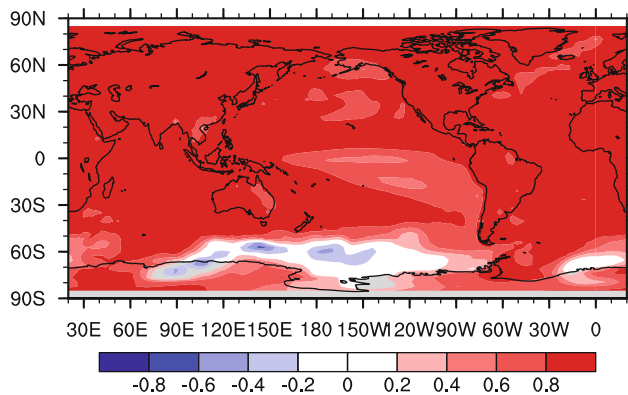
(c) Multi-model (HCST)   **(d)** Multi-model (HCST)
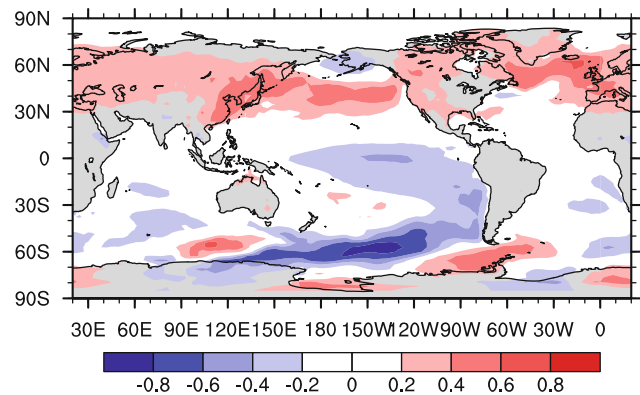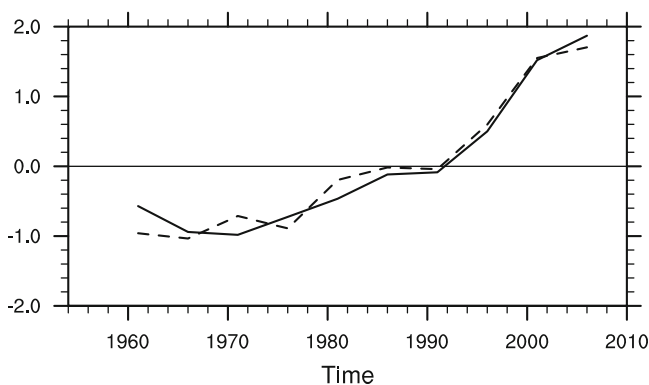
(e) Time series   **(f)** Time series
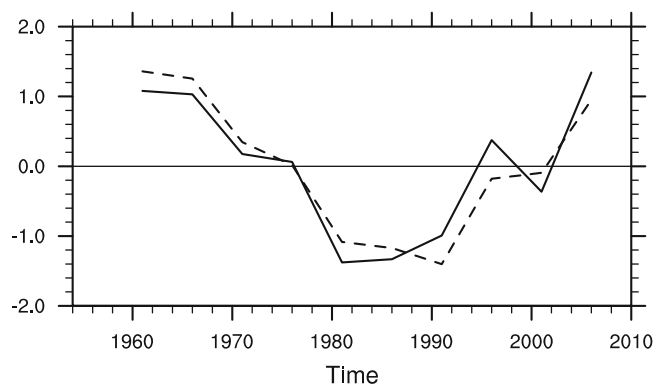
Fig. 2 First (*left*) and second (*right*) SVD modes of surface air temperature anomalies between the multi-model ensemble of the HCST experiment for a 1–3-year lead time and the NCEP reanalysis (Kalnay et al. 1996) for its corresponding time. The SVD analysis was performed in the global domain. The squared covariance fraction (SCF) explained by the SVD mode and the temporal correlation coefficient (R) between the expansion coefficients for the observation and multi-model ensemble fields are indicated above (**a**) and (**b**). *Top* and *middle panels* indicate homogeneous correlation maps of the observation and the multi-model ensemble, respectively. Bottom panels are time coefficients for (**e**) first and (**f**) second SVD modes in the observation (*broken line*) and the multi-model ensemble of the HCST experiments (*solid line*)

OBS (NCEP)

(a) Linear trend

NoAS

(b) Linear trend

(c) Regression on global mean temperature

(d) Regression on global mean temperature

(e) SVD1

(f) SVD1
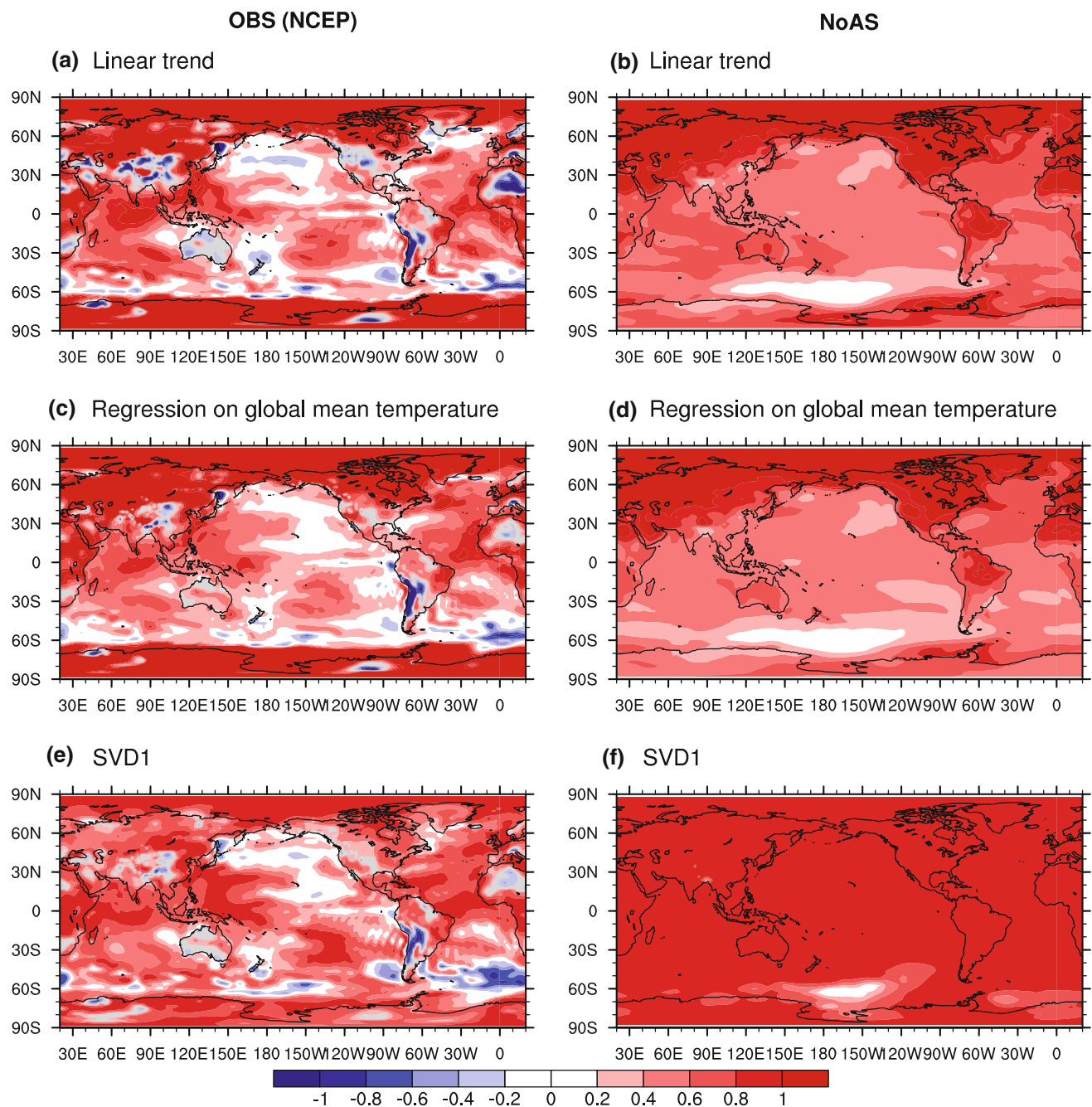
Fig. 3 Externally forced variations of surface air temperature anomalies estimated from **a, b** linear trend (°C/50-year), **c, d** regression maps with global mean temperature anomalies (°C/°C), and **e, f** first SVD mode between the NoAS experiment and observation (°C). Left and right panels are the observation (Kalnay et al. 1996) and the NoAS experiment, respectively. Linear trend and regression maps are calculated by annual anomalies over 1955–2010 period while SVD analysis are performed for 3-year-mean from 1961 to 2006 every 5 years (same as Fig. 2)

related to AMO (Delworth and Mann 2000; Knight et al. 2005; Zhang 2007). These higher predictive skills in VAT300 compared to those in SAT suggest that the initialization of ocean temperature variability contributes to the enhancement of the predictive skill on decadal timescales.

The predictable components on decadal timescales associated with internally generated variability are higher in the subsurface ocean than on the surface. Figure 5 shows predictable areas where 3-year-mean ocean temperature anomalies at the surface and the depths of 100 and 300 m are significantly hindcasted at 95 % confidence levels in

terms of ACC and the ratio of RMSE in the HCST experiment to that in the NoAS experiment is less than 0.9. The area with enhanced predictability of ocean temperature anomaly expands with the depth (Fig. 5). In particular, in the North Atlantic, predictability becomes prominent at 300-m depth but limited at the surface near the coast of eastern America and Greenland. Although ocean temperature variability near the sea surface is affected by high-frequency atmospheric variability, the long-term predictable memory associated with slow ocean variability would reside in the subsurface ocean.

Figure 6 shows an annual time series of ocean temperature anomalies at 300-m depth in the North Pacific, the South Pacific, and the North Atlantic. Observed ocean temperature anomalies in Ishii and Kimoto (2009) vary on decadal timescales (black lines in Fig. 6), while externally forced variability estimated from the multi-model ensemble in the NoAS experiment (red lines in Fig. 6) shows an almost monotonically increasing trend. In the North Pacific, the observed temperature anomalies display two positive phases for the 1960–1980 and 1990-2000 periods and two negative phases for the 1980–1990 and 2000–2010
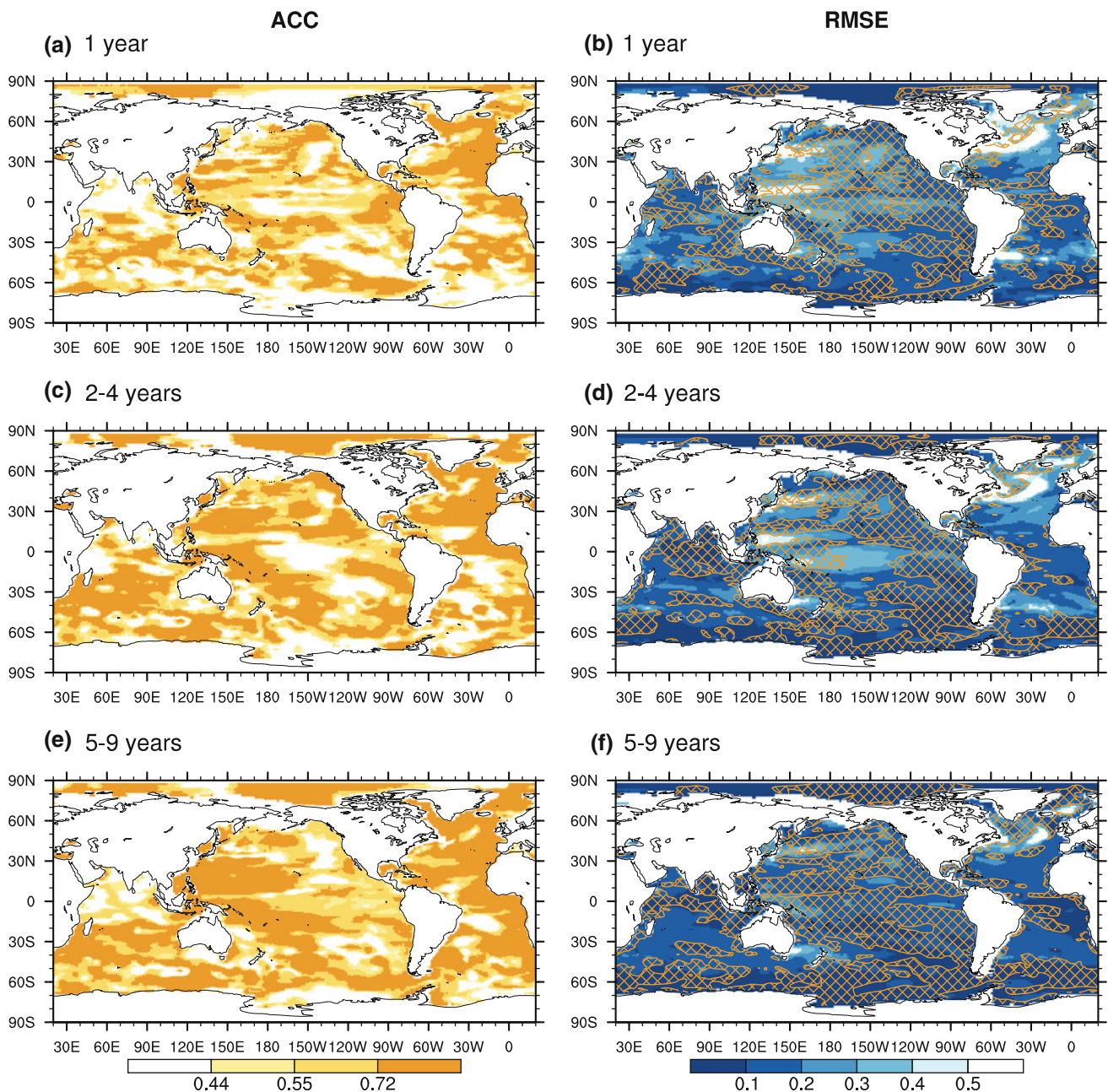
Fig. 4 Same as Fig. 1, but for vertically averaged temperatures from the ocean surface to 300-m depth

Predictable year (initialization impact)

(a) SST

(b) TO 100-m

(year)

(c) TO 300-m
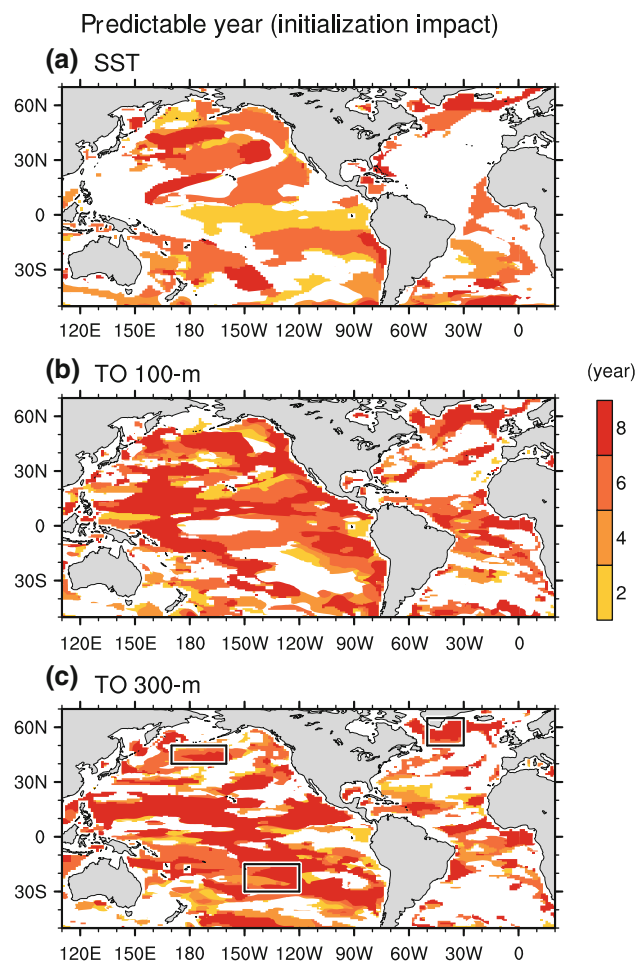
Fig. 5 Predictable years of the 3-year-mean ocean temperature anomalies at **a** the surface, **b** 100-m depth, and **c** 300-m depth in the HCST experiment. Shaded region indicates the areas with ACC higher than 0.55 (i.e., the statistically significance at the 95 % level) and RMSE of the HCST experiment to the NoAS less than 0.9. The bar colors indicate the best lead time for the predictability for each grid points and the lead time labeled by the middle year of 3-year-mean (e.g., a 2-year lead time indicates the mean of the 1–3 year lead time). *Black boxes* in (**c**) indicate regions in the North Pacific, the South Pacific, and the North Atlantic, as shown in Fig. 6

periods with an amplitude of 0.3 °C (Fig. 6a), which show the opposite phases of the PDO index, as shown in the next section. This observed temperature variability is well simulated in the HCST experiment (blue lines) and easily distinguishable from the warming trend represented by the NoAS experiment: ACCs for 10-year-mean temperature anomalies are 0.86 between the observation and HCST experiment and −0.19 between the observation and NoAS experiment. Similar results between the observation and the HCST and NoAS experiments were also obtained in the South Pacific (Fig. 6b), while a variance is about 1.2 times larger than that in the North Pacific (standard deviations of the observed temperature anomalies are 0.177 °C and 0.194 °C in the North and South Pacific, respectively). In

the North Atlantic, on the other hand, a temporal variation of observed temperature anomalies is significantly correlated with the time coefficient for the second SVD mode (Figs. 2f, 6c; ACC is 0.71) and the AMO index (ACC is 0.72), as shown in the next section. This temporal variation in the North Atlantic is observed with an overestimated intensity in the HCST experiment but not the NoAS experiment. Note that an abrupt temperature warming in the North Atlantic during mid 1990s in the observation is well simulated by the hindcast experiment initialized in 1996 (Fig. 6c). Recent studies (Robson 2010; Robson et al. 2012; Yeager et al. 2012) indicated that this abrupt warming is caused by changes in northward heat transport of the Atlantic Ocean associated with the Atlantic meridional overturning circulation (AMOC).

In summary, predictability on decadal timescales is prominent in the subsurface ocean rather than at the sea surface. Near the sea surface, externally forced components such as the global-warming signal are the main factors that explain decadal predictability, whereas the response patterns of SAT to external forcing are different between the observation and the model. On the other hand, initialization enhances the predictive skills of temperature variability, particularly in the subsurface ocean. Time series of temperature variability in the subsurface ocean show dominant natural variability such as PDO and AMO. In the next section, we focus on the predictability of the major climate variability: the global mean temperature, AMO, and PDO.

4 Internally generated and externally forced variabilities

As we have seen in the previous section, our HCST experiment, compared to the NoAS experiment, shows the improved predictive skill of temperature variability for the initial several years. When we focus on subsurface temperature variability, long-term predictability for a decade appears in the regions near the centers of action associated with PDO and AMO. In this section, we will show that the major decadal climate variability such as PDO and AMO is predictable for several years.

4.1 Global mean temperature

The global mean temperature trend is mainly explained by externally forced variability. Indeed, Fig. 7 shows a time series of globally averaged SAT in the observation (red line) and the HCST (blue) and NoAS experiments (green). On an annual timescale, the observed SAT gradually increases with interannual variations (top panels in Fig. 7). By applying longer time filtering such as a 3- and 5-year running mean, interannual variability becomes considerably
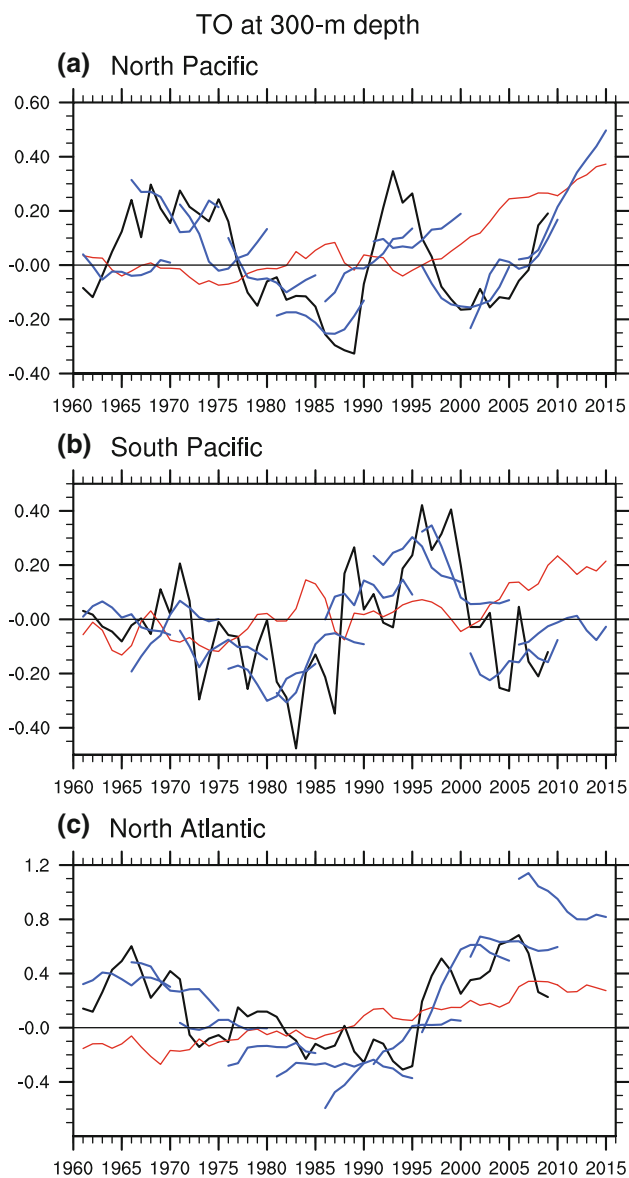
TO at 300-m depth

(a) North Pacific

(b) South Pacific

(c) North Atlantic

Fig. 6 Time series of annual ocean temperature anomalies at 300-m depth in **a** the North Pacific (40°N–50°N, 170°E–160°W), **b** the South Pacific (30°S–15°S, 150°W–120°W), and **c** the North Atlantic (50°N–65°N, 50°W–30°W). *Black*, *red*, and *blue lines* indicate the observation and the multi-model ensemble in the NoAS and the HCST experiments, respectively

smaller and then the temperature trend becomes prominent. This temperature trend is well simulated in both HCST and NoAS experiments (Fig. 7c–f) and is significantly correlated with the first SVD mode in Fig. 2 (ACC is 0.73).

The predictive skill of global mean temperature variability is slightly improved in the HCST experiment compared to the NoAS experiment although internally generated variability has a small impact on the globally averaged temperature variation. As a result of the small improvement by initialization, predictive skills in the multi-model ensemble, measured by ACC and RMSE, are

slightly higher in the HCST experiment than in the NoAS experiment (Fig. 7). In the recent decade, for example, the weakened warming trend in the observation is well simulated in the HCST experiment but not the NoAS experiment, particularly for the 2–4-year lead time (Fig. 7c, d). Mochizuki et al. (2010) suggested that a cool phase of PDO contributed to the weakened warming trend for the 2006–2008 period.

Statistical skill improvement of the global mean temperature by initialization is particularly exemplified by the RMSE. Figure 8 shows a histogram of the predictive skills estimated by an ensemble mean of 10 members obtained from random sampling of total 19 members in the three versions of MIROC. The probability density function (PDF) of RMSE in the HCST experiment (solid line) is smaller and narrower than that in the NoAS experiment (gray shaded area) for 1-, 2–4-, and 5–9-year lead times (bottom panels in Fig. 8), while the predictive skill measured by ACC is saturated for 2–4- and 5–9-year lead times (Fig. 8b, c). When the number of ensemble members is small, such as the three-member case (broke line), PDF of RMSE in the HCST experiment becomes broader than that in the NoAS experiment. Therefore, the predictive skill of the global mean temperature also depends on the number of ensemble members.

4.2 AMO

AMO is known as the multi-decadal SST variability in the North Atlantic (Delworth and Mann 2000; Murphy et al. 2010). Observational and modeling studies have indicated that AMO is related to rainfall variability in north east Brazil (Folland et al. 2001) and African Sahel regions (Folland et al. 1986; Rowell et al. 1995; Rowell 2003), summertime climate in North America and western Europe (Enfield et al. 2001; McCabe et al. 2004; Sutton and Hodson 2005), and Atlantic hurricanes (Goldenberg et al. 2001; Knight et al. 2006). For example, McCabe et al. (2004) suggested that recent droughts over the United States during 1999–2002 can be partly explained by a positive phase of AMO. This positive phase may have also affected the recent European summer heat wave during 2003 (Sutton and Hodson 2005). Although the mechanism of AMO is still controversial due to the insufficient historical observation in the deep ocean, study on AMO predictability is an important topic for climate prediction in terms of decadal timescales.

As shown in Figs. 2f and 6c, temperature anomalies of surface air and subsurface ocean in the North Atlantic display positive phases for the 1960–1970 and 2000–2010 periods and a negative phase for the 1980–1990 period. These temperature anomalies exhibit behavior similar to the AMO index (Enfield et al. 2001; Knight et al. 2005;

Fig. 7 Time series of globally averaged surface air temperature in the HCST and NoAS experiments (*left* and *right panels*, respectively) for **a** 1-, **c** 2–4-, and **e** 5–9-year lead times for their corresponding times. *Red, blue*, and *green colors* indicate the observation, and the HCST and NoAS experiments, respectively. *Cross, square*, and *triangle marks* indicate each ensemble member in MIROC3m, MIROC5, and MIROC4h, respectively. The correlation coefficient and RMSE are given for the multi-model ensemble

Global mean temperature
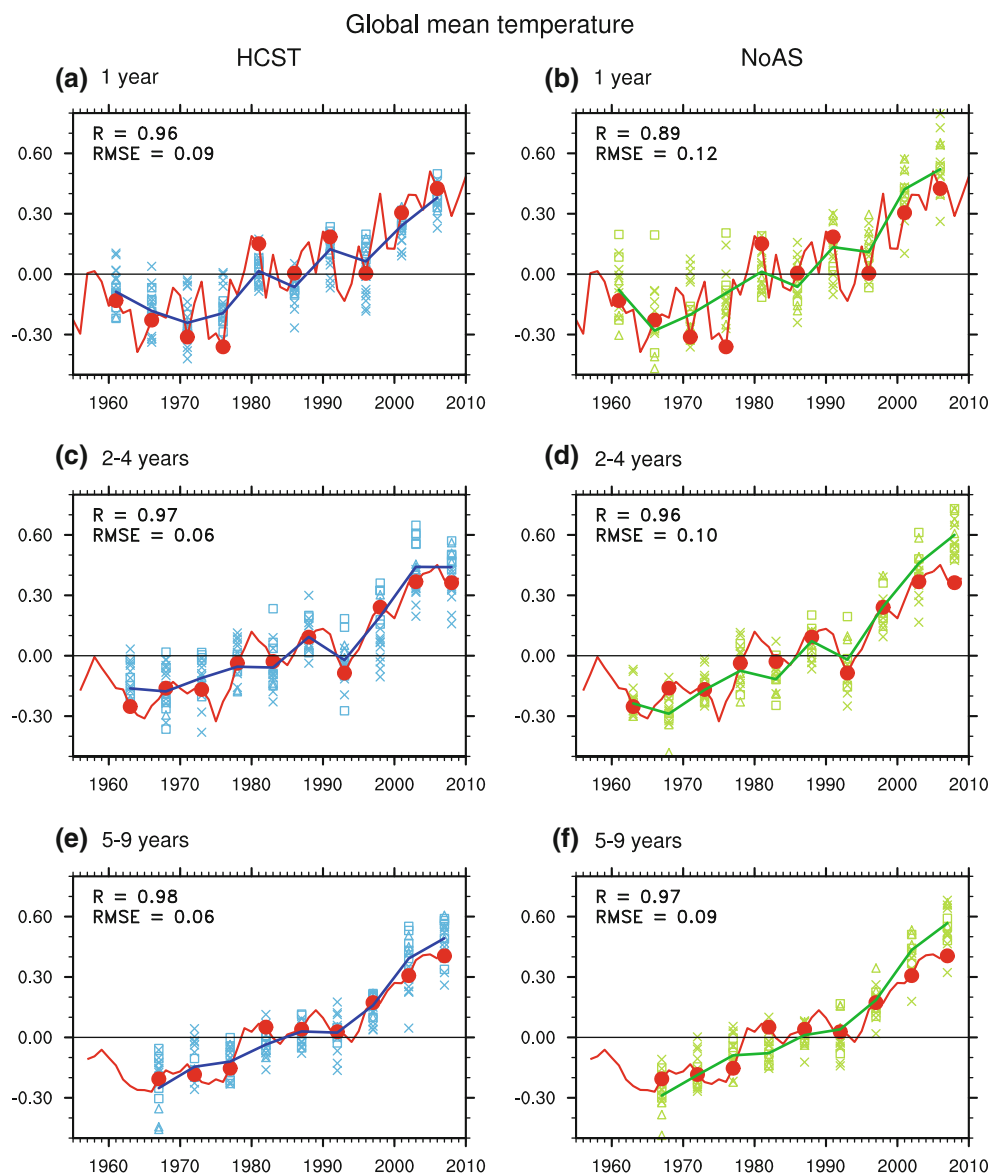
Sutton and Hodson 2005; Trenberth and Shea 2006). Figure 9 shows the time series of the AMO index proposed by Trenberth and Shea (2006); i.e., SST anomalies averaged over the North Atlantic (EQ–60°N, 80°W–0°W) minus global SST anomalies (60°S-60°N). For 1- and 2–4-year lead times, the observed temporal variation of the AMO index is well simulated in the HCST experiment (ACCs are 0.79 and 0.86, respectively) but is slightly underestimated in the multi-model ensemble (standard deviations at 1- and 2–4year lead times are 0.15 and 0.13 in the observation but 0.11 and 0.08 in the multi-model ensemble, respectively). Although the NoAS experiment also appears to simulate the AMO index involved in the warming trend, the predictive skills of the AMO index in the HCST experiment are higher than those in the NoAS experiment for the initial five years (Fig. 10). Consistent with these results, the improved

predictive skills in the HCST experiments appear in ACC and RMSE histograms based on 10-member ensembles (Fig. 11a, b, d, e). However, for the 5–9-year lead time, the simulated AMO indices in the HCST and NoAS experiments have similar histograms of prediction skills in particular in ACC (Fig. 11c, f), implying that the initialization impact on SST anomalies associated with AMO is effective for the initial five years.

Our HCST experiments, including a total of 19 members, facilitated the investigation of the predictive skills related to the number of ensemble members. Random sampling of 19 members in the three MIROC versions included not only initial uncertainty but also model uncertainty. A histogram of the skills with 10,000-times resampling shows a normal distribution, particularly in RMSEs (Fig. 11), implying that our bootstrap method has a
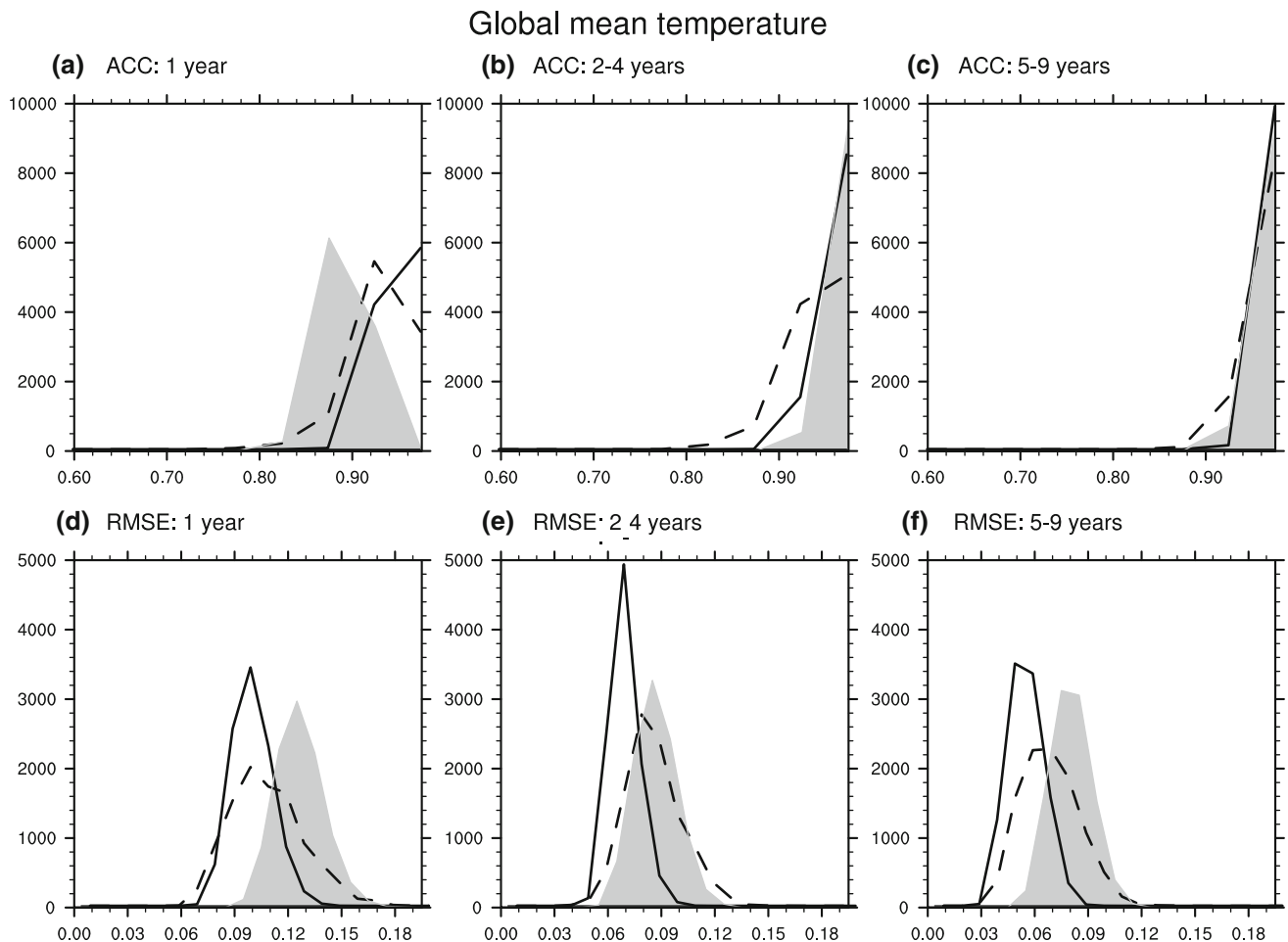
Fig. 8 Histogram of anomaly correlation coefficients (*top*) and RMSE (*bottom*) of globally averaged surface air temperature in the HCST experiment with 10,000 times random resampling of 19 members. *Solid* and *broken lines* indicate 10 and three members in the HCST experiments. *Gray shaded area* indicate the prediction skill by 10 members in the NoAS experiments

sufficient number of samples. Figure 12 shows the predictive skills of the AMO index in the HCST experiment as a function of the number of ensemble members. By increasing the number of ensemble members, the predictive skill becomes higher and its range reduces in both ACC and RMSE. The skill improvement of the median with the increase of ensemble members is prominent for less than 5 ensemble members, while the skill range tends to be saturated for more than 10 ensemble members taken from a mixture of all ensemble members in the three MIROC versions. Note that the ACC skill ranges of the AMO index have a long tail toward low predictive skills (Fig. 11a–c). These PDFs of ACC are affected by the upper limit of the correlation coefficient.

Although the predictive skill in MIROC3m is low for 2–4- and 5–9-year lead times, the multi-model ensemble of the three MIROC versions has a remarkable predictive skill. In particular, for 2–4-year lead time, the multi-model ensemble shows the best predictive skill as compared to each MIROC ensemble version (Fig. 12b, e). As shown in

Sect. 2, our HCST experiments have a different number of ensemble members among three MIROC versions. Due to the small number of ensemble members, MIROC4h shows unstable predictive skills of the AMO index (circle in Fig. 12): the lowest for the 1-year lead time, highest for the 2–4-year lead time, and medium for the 5–9-year lead time. Our higher predictability in the multi-model ensemble suggests a possibility that useful predictive skills will also be obtained by the CMIP5 multi-model ensemble in many research centers.

In our multi-model ensemble, the HCST experiment shows higher predictive skill of the AMO index than the NoAS experiment for the initial 5-year lead time but lower at 7-year and 5–9-year lead times (Figs. 9e, f, 10). The lower predictive skills of the HCST experiment for longer lead time may be related to large forecast uncertainty. From 4-year to 5-year lead times, predictive skills of the AMO index suddenly decrease in both of the HCST and NoAS experiments (Fig. 10). Associated with this decrease, magnitude of ACC error bar also increases in the

Fig. 9 Same as Fig. 7 but for the AMO index
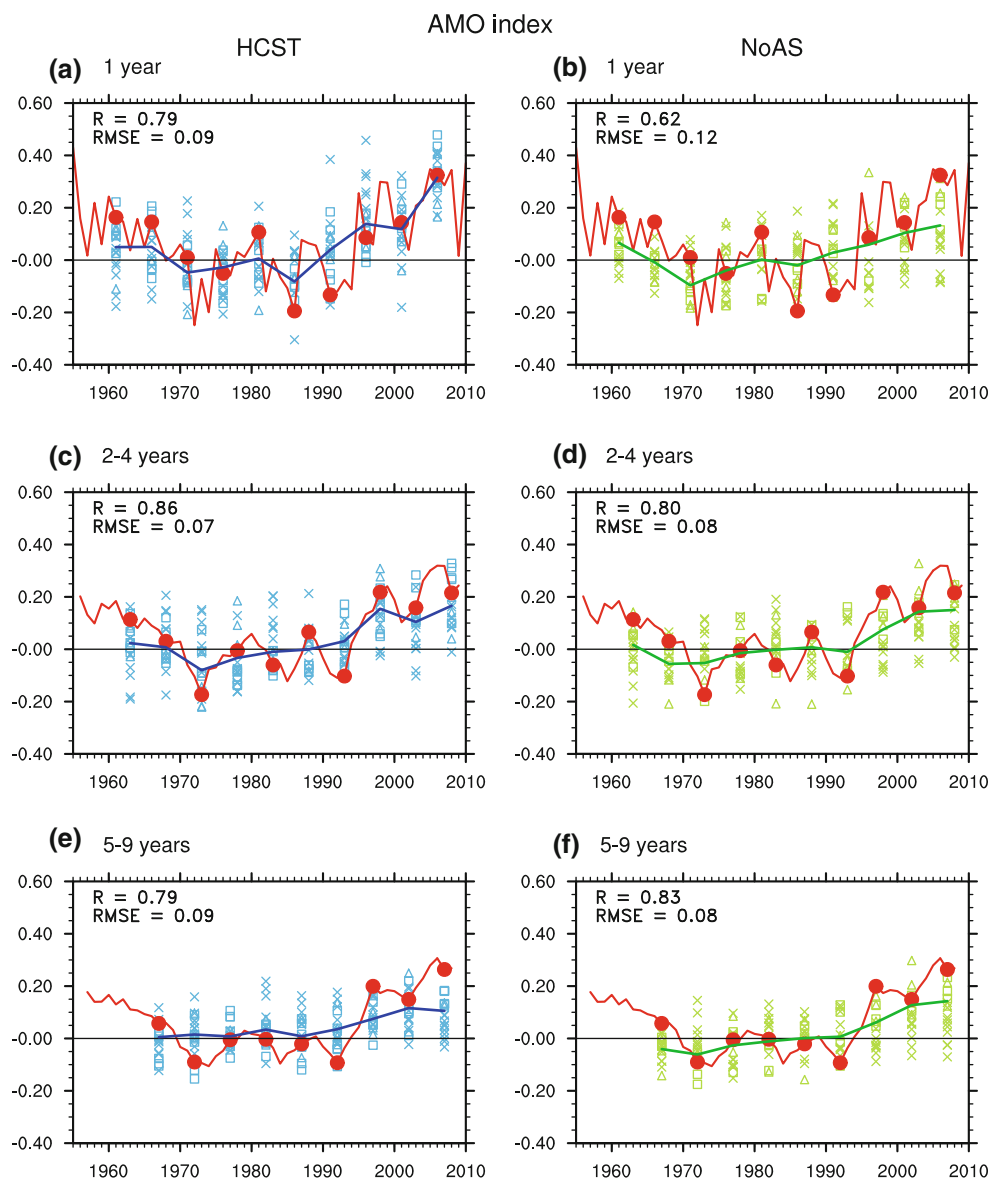
HCST experiment from 0.14 to 0.2 during this period. After 5-year lead time, the error bar in the HCST experiment becomes more than 2.0 while that in the NoAS experiment retains about 0.15. In addition to this large uncertainty in the HCST experiment, predictive skill of the AMO index suffers from a small sampling of hindcast because our analysis period covers only one phase of AMO. When we focus on the longer lead time such as 7-year or 5–9-year, the recent upward trend of the AMO index becomes prominent due to a limited sampling of the downward trend during 1960s (Fig. 9e, f). In our multi-model ensemble, however, the reason for the high predictive skill of AMO in the NoAS experiment is still unknown. To reveal predictability of AMO, a perfect model approach and a reproduction of climate for the past century are an important topic for future studies.

4.3 PDO

The PDO is the dominant climate variability in the North Pacific on decadal timescales (Mantua et al. 1997). During the positive phase of PDO, SST anomalies became cool in the central North Pacific and warm along the western coast of North America, and vice versa. Meehl et al. (2009) pointed out that the abrupt climate change observed in 1970s (known as the climate shift) contributed to decadal variability in PDO from negative to positive phases as well as externally forced variability. Mochizuki et al. (2010) indicated that the initialized ensemble hindcast in MIROC3m demonstrated a predictive skill of PDO for several years. In their follow-up paper (Mochizuki et al. 2012), reasonable predictability of PDO was also obtained by the initialized hindcasts in MIROC5 but not in MIROC4h due to the
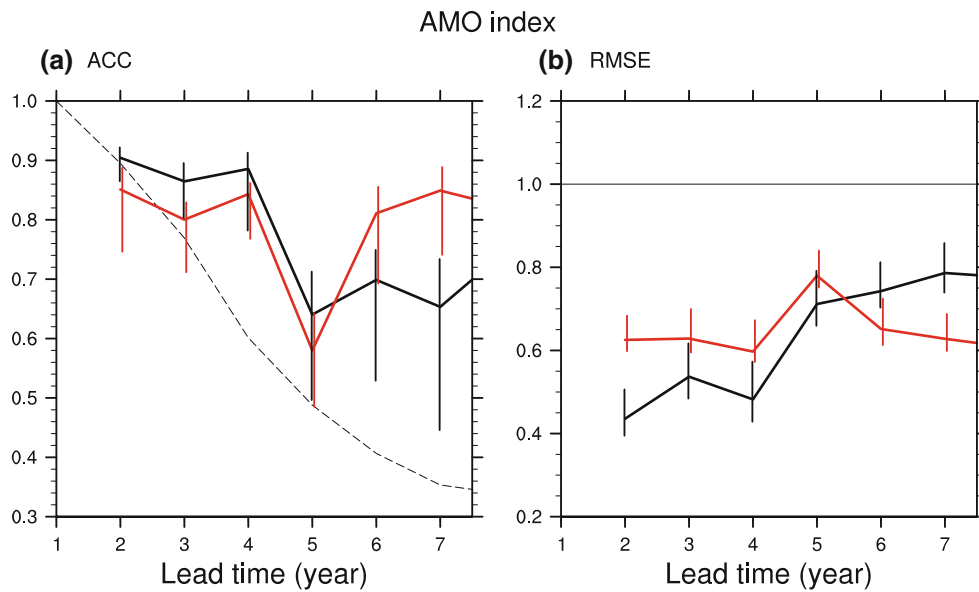
Fig. 10 Three-year-mean predictive skill of the AMO index in the HCST (*black*) and NoAS experiments (*red*), estimated by the multi-model ensemble. Predictive skills are measured by **a** anomaly correlation coefficient (ACC) and **b** root-mean-squared error (RMSE). *Broken line* indicates persistent prediction estimated from autocorrelation. *Error bars* indicate the range from 10 to 90 % of predictive skills obtained by multi-model ensemble, in which the ensemble mean in each model is estimated from random sampling repeated 10,000 times. RMSE is normalized by one standard deviation of the observed AMO index during the 1955–2010 period

small number of ensemble members. On the basis of these hindcast experiments with different skills of PDO, we focus on the impact of the multi-model ensemble on the predictability of PDO in this subsection.

To obtain a time series for the PDO index in the observation and the HCST and NoAS experiments, we used a common definition of PDO. In a manner similar to the definition in Mantua et al. (1997), we defined the spatial pattern of PDO as the leading empirical orthogonal function (EOF) of the observed and linearly detrended VAT300 anomaly (Ishii and Kimoto 2009) over the North Pacific (north of 20°N) during the 1955–2010 period. For this spatial pattern of PDO, time coefficients were obtained by the linearly detrended projection of VAT300 anomalies in the observation and the HCST and NoAS experiments. These linearly detrended time series are defined as the PDO index in this paper. As shown in Fig. 2a, c, the spatial patterns of temperature trend are not identical for the observation and the HCST experiment: SAT anomalies in the North Pacific have almost no trend in the observation but a warming trend in the HCST experiment. This discrepancy implies model deficiencies related to the response to external forcing.

Figure 13 shows a time series for the PDO index in the observation and the HCST and NoAS experiments for 1-, 2–4-, and 5–9-year lead times. The observed PDO index has a negative phase from 1960 to 1975, a positive phase in 1980s, and a negative phase again for the recent five years. This temporal variation having an opposite sign is also observed in the subsurface ocean, as shown in Fig. 6a, b. For 1- and 2–4-year lead times, the observed temporal variations of the PDO index are reproduced in the HCST experiment but not in the NoAS experiment (Fig. 13a–d). In particular, the positive phase during 1980s is observed in the HCST experiment although the amplitude in the multi-model ensemble is considerably smaller than that in the observation. However, for the 5–9-year lead time, temporal variations in the HCST experiment are almost the same as those in the NoAS experiment and different from observations, suggesting that in our present system, the PDO index is less predictable beyond the 5-year lead time.

Consistent with these time series, the 3-year-mean PDO index has the predictive skills for about 5 years in the HCST experiment but lower in the NoAS experiment (Fig. 14). For the initial 5-year lead time, the HCST experiment shows significant ACC values and RMSE values lower than the observed standard deviation (black lines in Fig. 14). On the other hand, the PDO index in the NoAS experiment is almost uncorrelated with the observation and has RMSE values higher than the observed standard deviation (red lines in Fig. 14). When a number of ensemble members is limited, such as for the 10-member case, ACC in the NoAS experiment covers a broader range of values compared with that in the HCST experiment (Fig. 15). The broad range of ACC values in the NoAS experiment implies that our PDO index has larger variance of internally generated variability than that of externally forced variation. Mochizuki et al. (2012) showed that the
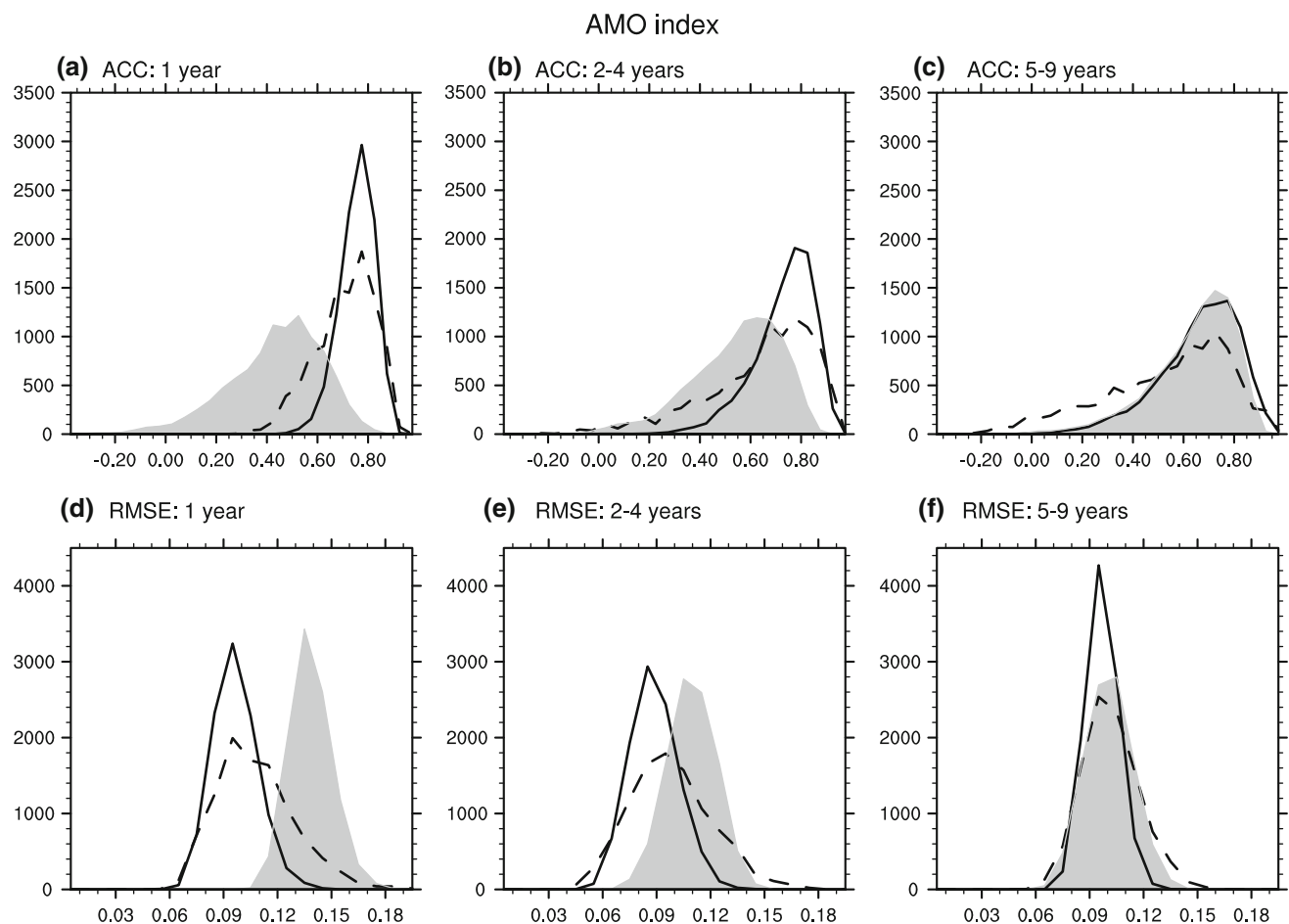
AMO index

Fig. 11 Same as Fig. 8 but for the AMO index

small number of ensemble members degrades the hindcast skills of PDO. Consistent with their results, the histogram of the HCST experiment with the 10-member ensemble is sharper than those of the HCST experiment with the three-member ensemble and the NoAS experiment with the 10-member ensemble (Fig. 15). These results indicate that the predictive skill of the PDO would be enhanced by increasing the number of ensemble members or the multi-model ensemble performed by many research centers.

Although we used only the three models in the HCST experiment, the stable predictive skill of the PDO index is obtained in our multi-model ensemble. Figure 16 shows the predictive skills of the PDO index associated with the number of ensemble members. By increasing the number of ensemble members, the averaged predictive skill tends to improve except the ACC values for the 5–9-year lead time. In particular, for 1- and 2–4-year lead times, the multi-model ensemble provides the predictive skills identical to or higher than the 19-member ensemble mean (asterisk and red circle in Fig. 16). These results suggest

that the multi-model ensemble is effective in improving the predictive skill on decadal timescales.

Previous studies have suggested that PDO can be regarded as the North Pacific expression of the basin-wide SST variability called the Interdecadal Pacific Oscillation (IPO) (Power et al. 1999; Folland et al. 2002; Meehl et al. 2009). In fact, subsurface temperature anomaly in the North Pacific is correlated to that in the South Pacific with a 4-year lag (Fig. 6a, b); ACC is 0.32 and exceeds the statistical significance at the 95 % level by using a two-side Student t-test with 50 degrees of freedom. However, the mechanism of PDO or IPO is still controversial, e.g., in terms of the origin of PDO or IPO, the tropics or extratropics (Liu 2012). In addition, the three MIROC versions have different characteristics of tropical SST variability; the amplitude of ENSO is high in MIROC5 but low in MIROC3m (Watanabe et al. 2010). To enhance the predictive skill on decadal timescales, physical processes contributing to interannual-to-decadal ENSO phenomena and tropical Pacific variability are an important topic for future studies.
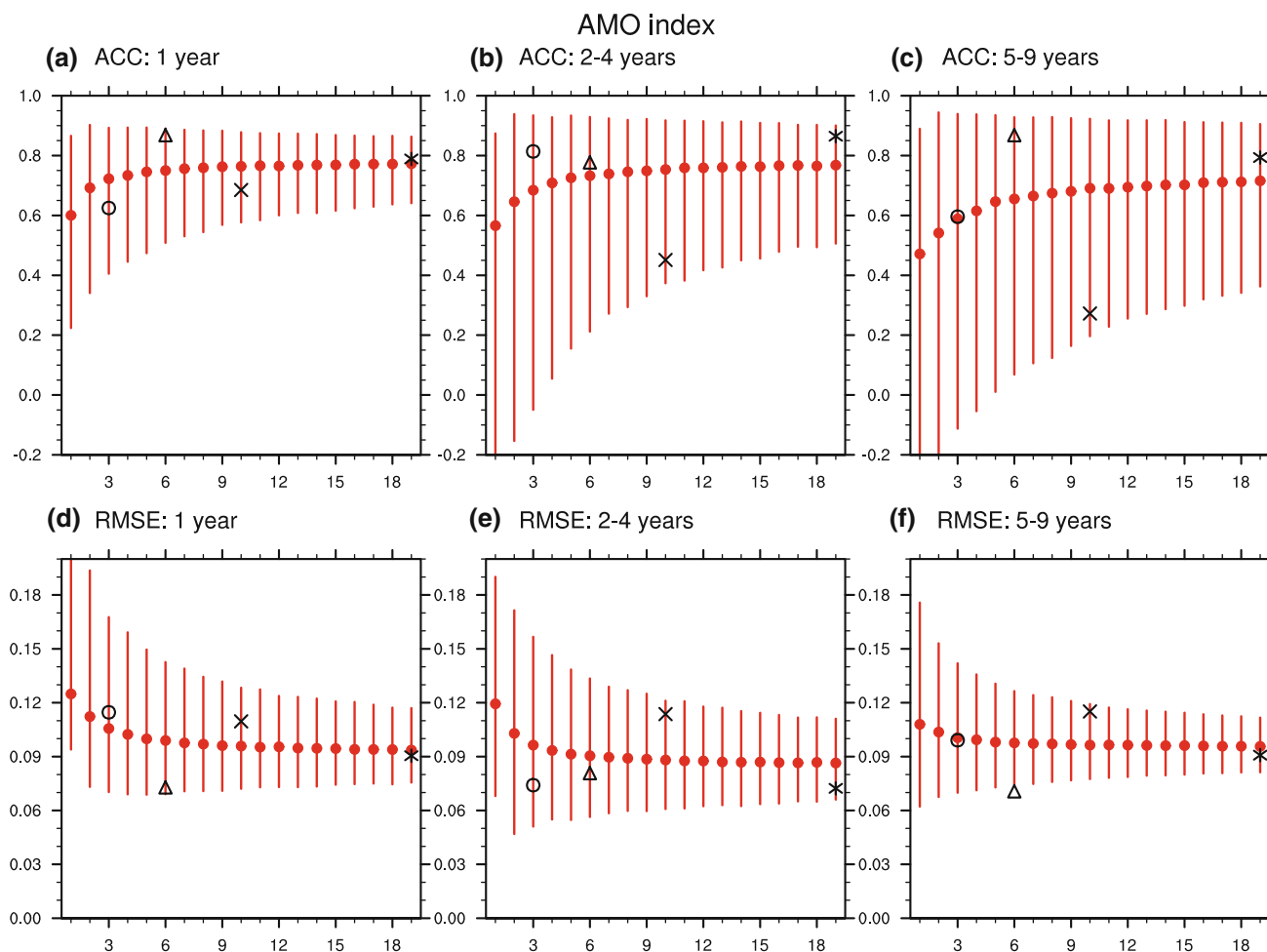
Fig. 12 Anomaly correlation coefficients (*top*) and RMSE (*bottom*) of the AMO index in the HCST experiment associated with the ensemble-averaged number of members. *Circles* and *bars* indicate the median and the range from 1 to 99 % of predictive skills, respectively, when random sampling was repeated 10,000 times for 19 members. *Cross*, *triangle*, *square*, and *asterisk marks* indicate ensemble means in MIROC3m, MIROC4h, MIROC5, and multi-model, respectively

5 Discussion and conclusions

We have investigated the predictability of decadal climate variability by a multi-model ensemble using three versions of MIROC. In our HCST experiments, initial conditions were obtained in a similar manner on the basis of anomaly assimilation in each model. We used the same external forcing to drive the initialized and uninitialized runs for each model. Therefore, uncertainties due to both initial errors and model imperfection are contained in the HCST ensemble, which can be used to assess the initialization impacts by comparing the initialized and uninitialized runs with sufficient ensemble size.

Our multi-model ensemble shows that the most predictable component in SAT variability originates in external condition (Figs. 1, 2, 7) and that internally generated variability is of secondary importance in SAT variability on decadal timescales. However, this does not

imply that internal variability targeted by initialization is negligible in near-term climate prediction. For example, with respect to the global mean temperature variability, the impact of initialization has not been considerably clear in the case of the twentieth century but has been significant in the case of the recent decade (Fig. 7c, d); RMSEs of global mean temperature anomalies at 2–4-year lead time in the HCST and NoAS experiments are 0.05 and 0.06 by excluding hindcasts initialized on 2001 and 2006 (not shown), but 0.06 and 0.10 by including these initialized hindcasts, respectively. Furthermore, when we focus on the subsurface ocean, the observed interannual-to-decadal variability is large compared to the linear trend or the externally forced variability estimated by the NoAS experiment, which is consistent with long-term predictability related to AMO and PDO signals (Figs. 4, 5, 6). Our results suggest that the long-term predictable memory related to AMO and PDO resides in the subsurface ocean.
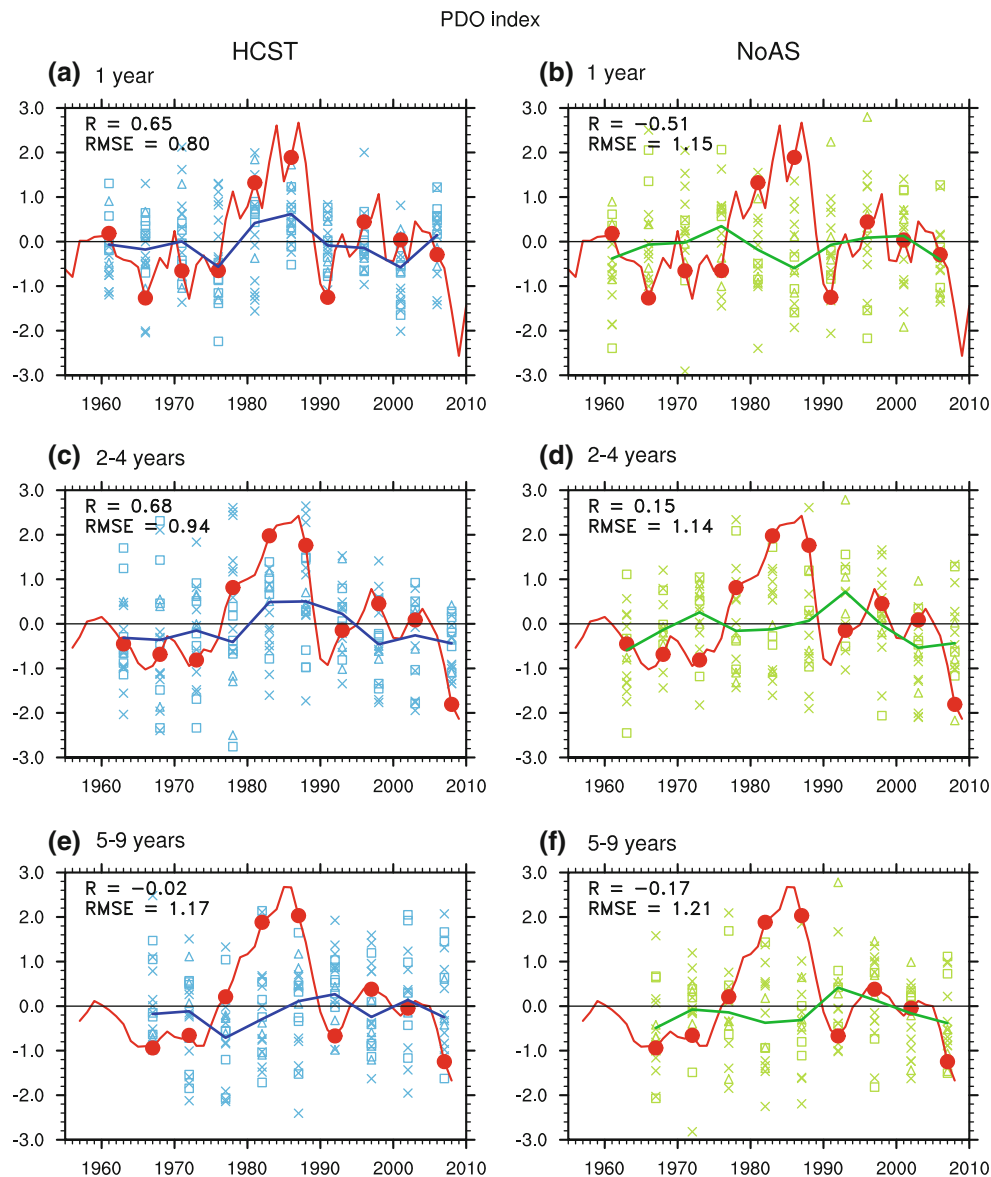
Near the sea surface, ocean temperature variability is strongly affected by unpredictable atmospheric high-frequency noise and externally forced variability. These variations may reduce the predictive skills of SAT variability associated with AMO and PDO.

In our current prediction system, AMO and PDO indices have the predictive skills for several years (Figs. 10, 12, 14, 16) although the mechanisms of these phenomena are still unknown. The SVD analysis indicates that a predictable component of surface temperature anomalies shows a mixture of PDO-like and AMO-like patterns (Fig. 2). Some recent modeling and observational studies suggest that a low-frequency component of PDO on multi-decadal time-scales is responsible for AMO with a lag of about a decade (Zhang and Delworth 2007; d'Orgeville and Peltier 2007). Possible physical processes involved in the relationship

between PDO and AMO may be clarified by further analyzing the initialized HCST experiments, which will be a topic of future work.

To contribute to the next report of IPCC AR5, hindcast and forecast experiments will be conducted using many CMIP models with different ensemble sizes. Our multi-model ensemble shows higher predictive skills in the dominant ocean temperature variability such as AMO and PDO compared to any of the single-model ensemble. This suggests that the multi-model ensemble approach in CMIP5 also has the potential to increase the predictability by reducing errors in single models. Further studies that unravel physical processes of the predictable natural variability are desirable in order to enhance our understanding of these processes and increase the predictive skills of decadal climate variability.
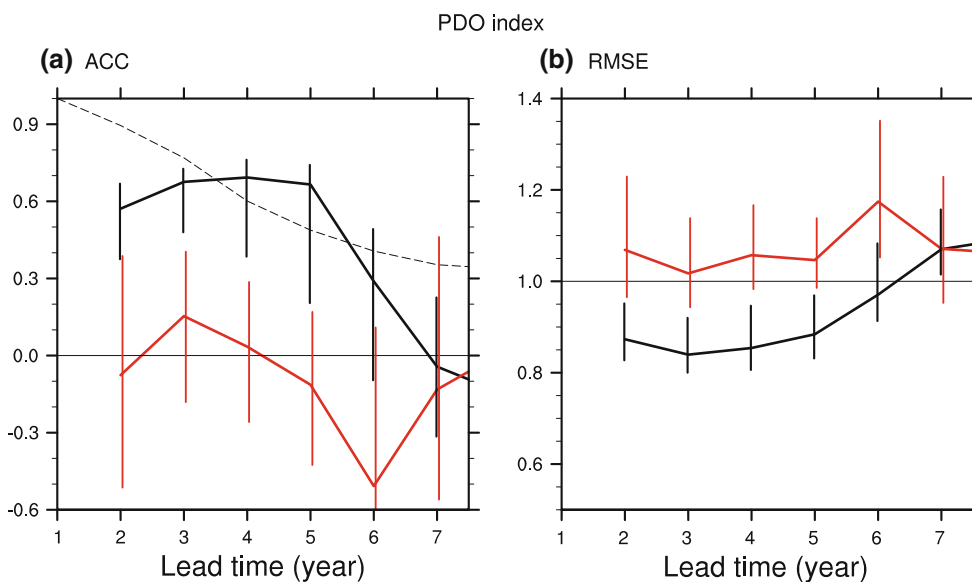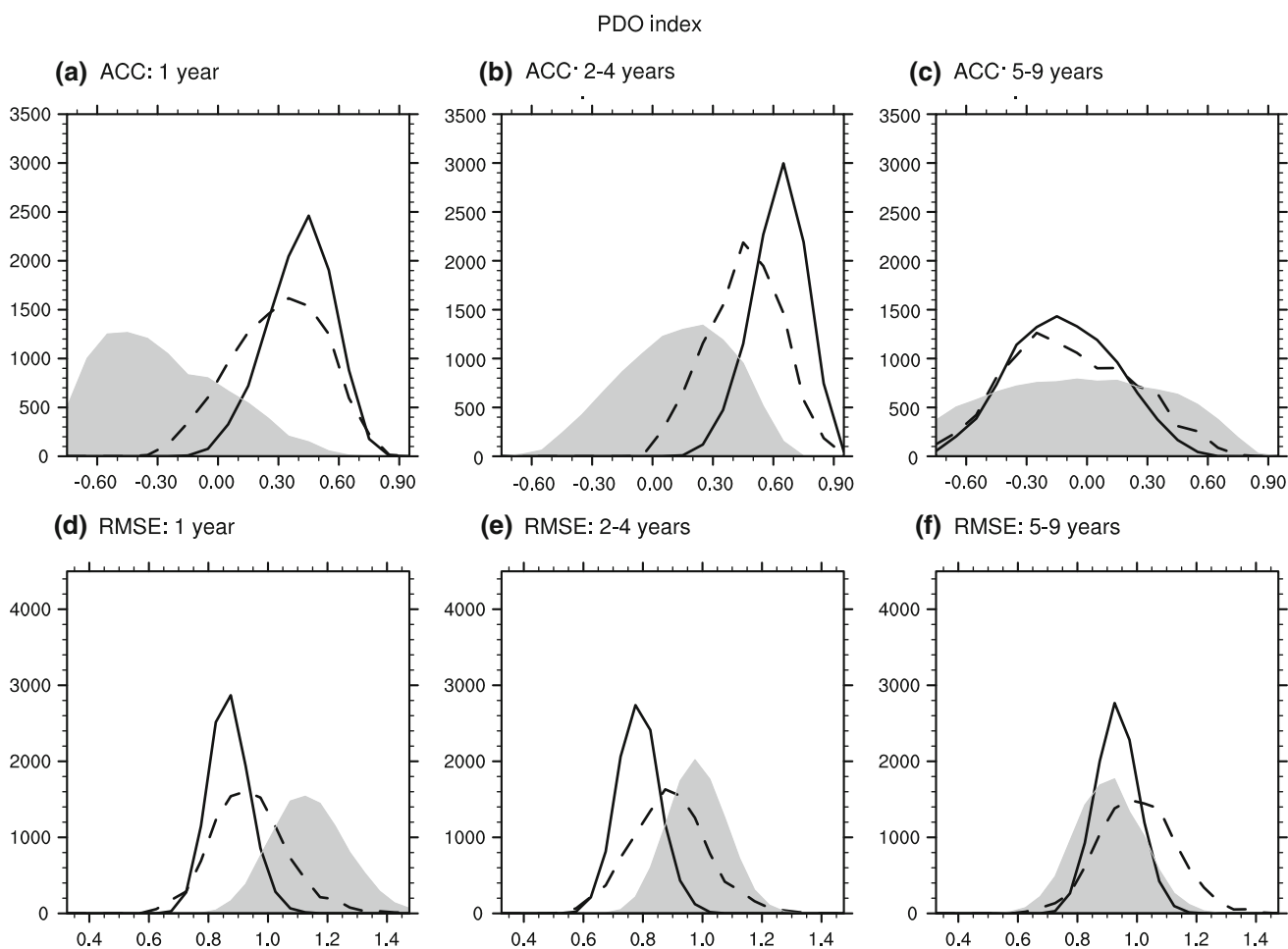
PDO index

Fig. 14 Same as Fig. 10 but for the PDO index

PDO index

Fig. 15 Same as Fig. 8 but for the PDO index
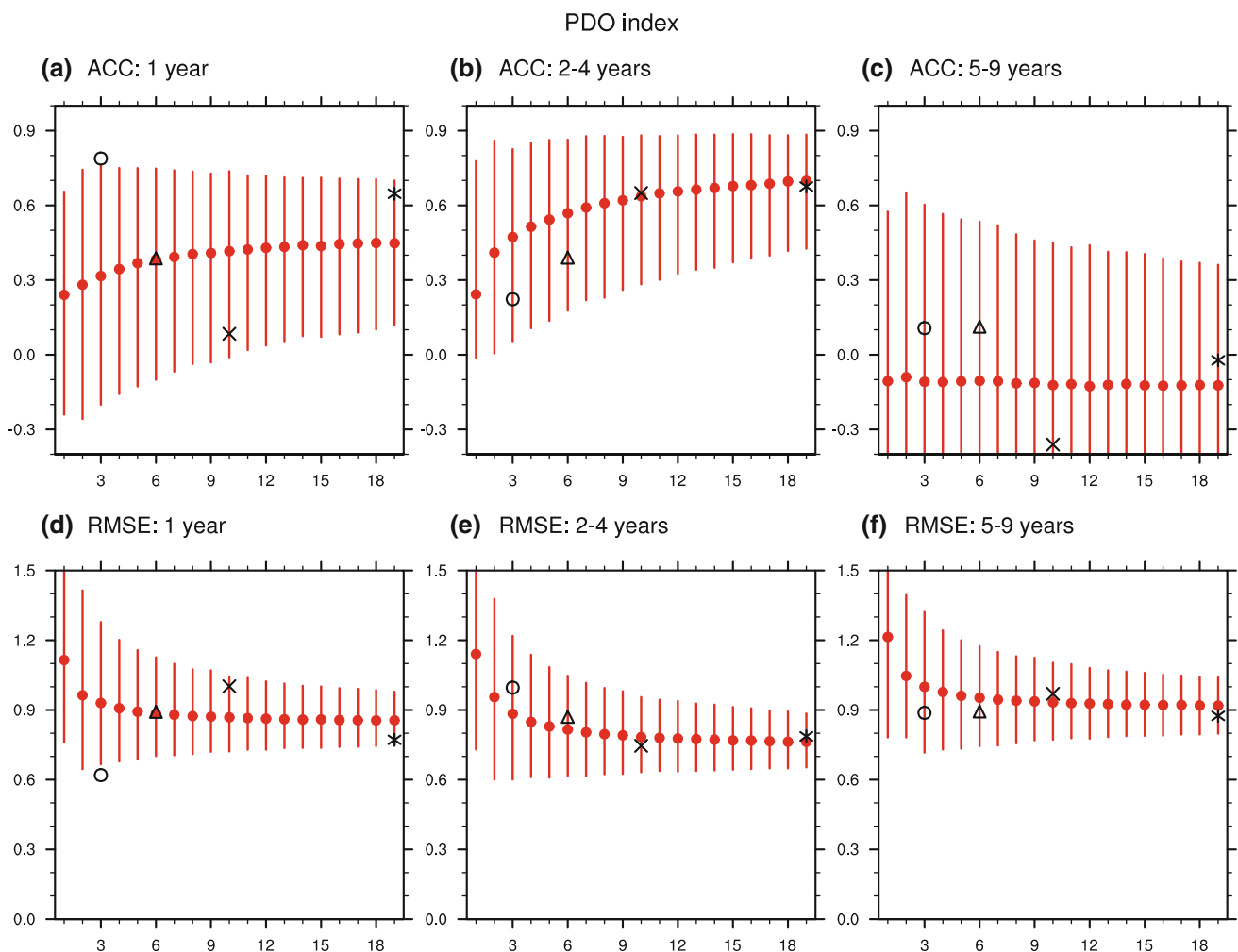
PDO index

Fig. 16 Same as Fig. 12 but for the PDO index

References

Bloom SC, Takacs L, da Silva AM, Ledvina D (1996) Data assimilation using incremental analysis updates. Mon Weather Rev 124:1256–1271

Bretherton CS, Smith C, Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. J Clim 5:541–559

Chikamoto Y, Kimoto M, Ishii M, Watanabe M, Nozawa T, Mochizuki T, Tatebe H, Sakamoto TT, Komuro Y, Shiogama H, Mori M, Yasunaka S, Imada Y, Koyama H, Nozu M, Jin F (2012) Predictability of a stepwise shift in Pacific climate during the late 1990s in hindcast experiments using MIROC. J Meteorol Soc Jpn (special issue, in press)

Cox P, Stephenson D (2007) A changing climate for prediction. Science 317:207–208. doi:10.1126/science.1145956

Delworth TL, Mann ME (2000) Observed and simulated multidecadal variability in the northern hemisphere. Clim Dyn 16(9):661–676

Delworth TL, Rosati A, Anderson W, Aderoft AJ, Balaji V, Benson R, Dixon K, Griffies SM, Lee HC, Pacanowski RC, Vecchi GA, Wittenberg AT, Zeng F, Zhang R (2011) Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. J Clim (in press)

Doblas-Reyes F, Weisheimer A, Palmer T, Murphy J, Smith D (2010) Forecast quality assessment of the ensembles seasonal-to-decadal stream 2 hindcasts. Tech. rep., Tech Memo 619, ECMWF, http://www.ecmwf.int/publications

d'Orgeville M, Peltier WR (2007) On the Pacific decadal oscillation and the Atlantic multidecadal oscillation: might they be related. Geophys Res Lett 34(23):L23,705

Enfield DB, Mestas-Nunez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relationship to rainfall and river flows in the continental US. Geophys Res Lett 28:2077–2080

Folland CK, Palmer TN, Parker DE (1986) Sahel rainfall and worldwide sea temperatures, 1901–85. Nature 320:602–607

Folland CK, Colman AW, Rowell DP, Davey MK (2001) Predictability of northeast Brazil rainfall and real-time forecast skill, 1987–98. J Clim 14(9):1937–1958

Folland CK, Renwick JA, Salinger MJ, Mullan AB (2002) Relative influences of the interdecadal pacific oscillation and enso on the south pacific convergence zone. Geophys Res Lett 29:21. doi:10.1029/2002GL014201

Goldenberg SB, Landsea CW, Mestas-Nuñez AM, Gray WM (2001) The recent increase in Atlantic hurricane activity: causes and implications. Science 293(5529):474

Hasumi H (2006) CCSR Ocean Component Model (COCO), Version 4.0. Center for Climate System Research Rep 25 p 103 pp, [Avaiable online at http://www.ccsr.u-tokyo.ac.jp/hasumi/COCO/coco4.pdf]

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteorol Soc 90(8):1095–1107

Hibbard KA, Meehl GA, Cox PM, Friedlingstein P (2007) A strategy for climate change stabilization experiments. EOS 88:217–219

Huang B, Kinter J, Schopf P (2002) Ocean data assimilation using intermittent analyses and continuous model error correction. Adv Atmos Sci 19(6):965–992

INTERNATIONAL CLIVAR PROJECT OFFICE (2011) Decadal and bias correction for decadal climate predictions. International CLIVAR Project Office 150:CLIVAR Publication Series (not peer reviewed)

Ishii M, Kimoto M (2009) Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. J Oceanogr 65(3):287–299

Ishii M, Kimoto M, Sakamoto K, Iwasaki S (2006) Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. J Oceanogr 62:155–170

Jin E, Kinter J, Wang B, Park C, Kang I, Kirtman B, Kug J, Kumar A, Luo J, Schemm J et al (2008) Current status of ENSO prediction skill in coupled ocean–atmosphere models. Clim Dyn 31(6):647–664. doi:10.1007/s00382-008-0397-3

K-1 Model Developers (2004) K-1 coupled GCM (MIROC) description, K-1 Tech. Rep., vol 1. Cent. for Clim. Syst. Res., Univ. of Tokyo, Tokyo, 34 pp

Kalnay E, Kanamitsu M, Kistler R, Deaven WCD, Gandin L, Iredell M, Saha S, White G, Wollen J, Zhu Y, Ebisuzaki MCW, Higgins W, Janowiak J, Ropelewski KCMC, Wang J, Leetmaa A, Jenne RRR, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–471

Keenlyside NS, Latif M, Jungclaus J, Kornblueh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the north atlantic sector. Nature 453:84–88. doi:10.1038/nature06921

Knight J, Allan R, Folland C, Vellinga M, Mann M (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. Geophys Res Lett 32:L20,708

Knight J, Folland C, Scaife A (2006) Climate impacts of the Atlantic multidecadal oscillation. Geophys Res Lett 33:L17,706

Komuro Y, Suzuki T, Sakamoto TT, Hasumi H, Ishii M, Watanabe M, Nozawa T, Yokohata T, Nishimura T, Ogochi K, Emori S, Kimoto M (2012) Sea-ice in twentieth-century simulations by new miroc coupled models: a comparison between models with high resolution and with ice thickness distribution. J Meteorol Soc Jpn (in press)

Liu Z (2012) Dynamics of interdecadal climate variability: a historical perspective. J Clim (in press)

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. Bull Am Meteorol Soc 78:1069–1079

McCabe GJ, Palecki MA, Betancourt JL (2004) Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. Proc Natl Acad Sci USA 101(12):4136

Meehl GA, Hu A, Santer BD (2009) The mid-1970s climate shift in the pacific and the relative roles of forced versus inherent decadal variability. J Clim 22(3):780–792

Mochizuki T, Ishii M, Kimoto M, Chikamoto Y, Watanabe M, Nozawa T, Sakamoto TT, Shiogama H, Awaji T, Suiura N, Toyoda T, Yasunaka S, Tatebe H, Mori M (2010) Pacific decadal oscillation hindcasts relevant to near-term climate prediction. Proc Natl Acad Sci USA 107:1833. doi:10.1073/pnas.0906531107

Mochizuki T, Chikamoto Y, Kimoto M, Ishii M, Tatebe H, Komuro Y, Sakamoto TT, Watanabe M, Mori M (2012) Decadal prediction using a recent series of MIROC global climate models. J Meteorol Soc Jpn (special issue, in press)

Moss R, Babiker M, Brinkman S, Calvo E, Carter T, Edmonds J, Elgizouli I, Emori S, Erda L, Hibbard K, Jones R, Kainuma M, Kelleher J, Lamarque JF, Manning M, Matthews B, Meehl J, Meyer L, Mitchell J, Nakicenovic N, O'Neill B, Pichs R, Riahi K, Rose S, Runci P, Stouffer R, van Vuuren D, Weyant J, Wilbanks T, van Ypersele JP, Zurek M (2008) Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies. Tech. rep., Pacific Northwest National Laboratory (PNNL), Richland

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010) The next generation of scenarios for climate change research and assessment. Nature 463(7282):747–756

Murphy J, Kattsov V, Keenlyside N, Kimoto M, Meehl G, Mehta V, Pohlmann H, Scaife A, Smith D (2010) Towards prediction of decadal climate variability and change. Procedia Environ Sci 1:287–304

Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S, Gregory K, Grubler A, Jung T, Kram T et al (2000) Emissions scenarios, a special report of working group III of the intergovernmental panel on climate change. Cambridge University Press, New York

Nozawa T, Nagashima T, Shiogama H, Crooks SA (2005) Detecting natural influence on surface air temperature change in the early twentieth century. Geophys Res Lett 32:L20,719. doi:10.1029/2005GL023540

Nozawa T, Nagashima T, Ogura T, Yokohata T, Okada N, Shiogama H (2007) Climate change simulations with a coupled ocean-atmosphere GCM called the model for interdisciplinary research on climate: MIROC. CGER Supercomput Monogr Rep 12, cent. for Global Environ. Res., Natl. Inst. for Environ. Stud., Tsukuba.

Pohlmann H, Jungclaus JH, Köhl A, Stammer D, Marotzke J (2009) Initializing decadal climate predictions with the GECCO oceanic synthesis: effects on the North Atlantic. J Clim 22(14):3926–3938. doi:10.1175/2009JCLI2535.1

Power S, Casey T, Folland C, Colman A, Mehta V (1999) Interdecadal modulation of the impact of enso on australia. Clim Dyn 15(5):319–324

Robson JI (2010) Understanding the performance of a decadal prediction system. PhD thesis, University of Reading, UK

Robson J, Sutton R, Lohmann K, Smith D, Palmer MD (2012) Causes of the rapid warming of the North Atlantic Ocean in the mid 1990s. J Clim (in press)

Rowell DP (2003) The impact of Mediterranean SSTs on the Sahelian rainfall season. J Clim 16(5):849–862

Rowell D, Folland C, Maskell K, Ward M (1995) Variability of summer rainfall over tropical north Africa (1906–92): observations and modelling. Q J R Meteorol Soc 121(523):669–704

Sakamoto TT, Komuro Y, Ishii M, Tatebe H, Shiogama H, Hasegawa A, Toyoda T, Mori M, Suzuki T, Imada-Kanamaru Y, Nozawa T, Takata K, Mochizuki T, Ogochi K, Nishimura T, Emori S, Hasumi H, Kimoto M (2012) MIROC4h—a new high-resolution atmosphere-ocean coupled general circulation model. J Meteorol Soc Jpn (in press)

Shiogama H, Nozawa T, Emori S (2007) Robustness of climate change signals in near term predictions up to the year 2030: changes in the frequency of temperature extremes. Geophys Res Lett 34(12):L12,714

Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming decade from a global climate model. Science 317:796–799. doi:10.1126/science.1139540

Solomon A, Goddard L, Kumar A, Carton J, Deser C, Fukumori I, Greene A, Hegerl G, Kirtman B, Kushnir Y et al (2010) Distinguishing the roles of natural and anthropogenically forced decadal climate variability: implications for prediction. Bull Am Meteorol Soc 92:141–156

Sutton RT, Hodson DLR (2005) Atlantic ocean forcing of North American and European summer climate. Science 309(5731):115

Takata K, Emori S, Watanabe T (2003) Development of the minimal advanced treatments of surface interaction and runoff. Glob Planet Change 38(1–2):209–222

Tatebe H, Ishii M, Mochizuki T, Chikamoto Y, Sakamoto TT, Komuro Y, Mori M, Yasunaka S, Watanabe M, Ogochi K, Suzuki T, Nishimura T, Kimoto M (2012) Initialization of the climate model MIROC for decadal prediction with hydographic data assimilation. J Meteorol Soc Jpn (special issue, in press)

Taylor K, Stouffer R, Meehl G (2009) A summary of the cmip5 experiment design. http://cmipllnlgov/cmip5/docs/TaylorCMIP5design.pdf. Last accessed 26 July 2010

Trenberth KE, Hurrell JW (1994) Decadal atmosphere-ocean variations in the pacific. Clim Dyn 9:303–319

Trenberth KE, Shea DJ (2006) Atlantic hurricanes and natural variability in 2005. Geophys Res Lett 33(12):L12,704

van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. Clim Dyn 38:1263–1280

Wallace JM, Smith C, Bretherton CS (1992) Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. J Clim 5:561–576

Watanabe M, Suzuki T, Oishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T, Hasumi H, Tatebe H, Kimoto M (2010) Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J Clim 23(23):6312–6335. doi:10.1175/2010JCLI3679.1

Yeager S, Karspeck A, Danabasoglu G, Tribbia J, Teng H (2012) A decadal prediction case study: late 20th centruy North Atlantic ocean heat content. J Clim (in press)

Zhang R (2007) Anticorrelated multidecadal variations between surface and subsurface tropical north atlantic. Geophys Res Lett 34:L12,713

Zhang R, Delworth TL (2007) Impact of the atlantic multidecadal oscillation on north pacific climate variability. Geophys Res Lett 34(23):L23,708