Seasonal to Decadal Predictions With MIROC6: Description and Basic Evaluation

Takahito Kataoka1, Hiroaki Tatebe1, Hiroshi Koyama1, Takashi Mochizuki1,2, Koji Ogochi1, Hiroaki Naoe3, Yukiko Imada1, Hideo Shiogama4,5, Masahide Kimoto5, and Masahiro Watanabe5

1Research Center for Environmental Modeling and Application, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, 2Department of Earth and Planetary Sciences, Kyushu University, Fukuoka, Japan, 3Meteorological Research Institute, Tsukuba, Japan, 4National Institute for Environmental Studies, Tsukuba, Japan, 5Atmosphere and Ocean Research Institute, The University of Tokyo Kashiwa Campus, Kashiwa, Japan

Abstract The present paper presents results of seasonal-to-decadal climate predictions based on a coupled climate model called the Model for Interdisciplinary Research on Climate version 6 (MIROC6) contributing to the Coupled Model Intercomparison Project Phase 6 (CMIP6). MIROC6 is initialized every year for 1960–2018 by assimilating observed ocean temperature and salinity anomalies and full fields of sea ice concentration and by prescribing atmospheric initial states from reanalysis data. The impacts of updating the system on prediction skill are then evaluated by comparing hindcast experiments between the MIROC6 prediction system and a previous system based on MIROC version 5 (MIROC5). Skill of seasonal prediction is overall improved in association with representation and initialization of El Niño/Southern Oscillation (ENSO), the Quasi-Biennial Oscillation (QBO), and the Northern Hemisphere sea ice concentration in MIROC6. In particular, the QBO is skillfully predicted up to 3 years ahead with a maximum anomaly correlation exceeding \( r = 0.8 \). The prediction skill for the North Atlantic Oscillation in winter is also enhanced, but the prediction still suffers from model's inherent errors. On decadal timescales, MIROC6 has a larger fraction of areas of the globe with better surface temperature skill at all lead times than MIROC5, and it has predictive skill in the annual-mean sea surface temperature (SST) in the North Atlantic and the Pacific. In particular, MIROC6 hindcasts at 2–5 years lead time are able to capture the spatial structure of SST changes in the North Pacific and the eastern tropical Pacific associated with the 1970s regime shift better than MIROC5 hindcasts.

Plain Language Summary The present paper presents results of seasonal-to-decadal climate predictions based on a coupled climate model called the Model for Interdisciplinary Research on Climate version 6 (MIROC6) contributing to the Coupled Model Intercomparison Project Phase 6 (CMIP6). The MIROC6 prediction system takes into account observed ocean, sea ice, and atmospheric conditions in the initialization, whereas its predecessor (MIROC5) considers observed ocean conditions only. Experiments with MIROC6 without initialization are also carried out in order to quantify impacts of initialization. Prediction skill of MIROC6 new prediction system several months to a year ahead is improved over MIROC5. In particular, the tropical variability in the stratosphere, which could be a key for predictions of the North American and European winter climate, is well predicted beyond a year. Also, while the overall improvement over MIROC5 is modest, MIROC6 has substantial skill for predictions of annual-mean sea surface temperatures, particularly over the North Atlantic and the Pacific several years in advance. MIROC6 is able to better predict the variability in the North Pacific and the eastern tropical Pacific than initialized MIROC5 and uninitialized MIROC6 experiments, highlighting the importance of the system updates and observed information.

1. Introduction

The climate fluctuates on various timescales and in various patterns, such as the North Atlantic Oscillation (NAO), El Niño/Southern Oscillation (ENSO), and the Pacific decadal variability (PDV), giving rise to extreme events over the globe (e.g., Camargo & Sobel, 2005; Delworth et al., 2016; Kosaka & Xie, 2013). Also, short-term changes associated with the internal fluctuations of the climate system can offset or exacerbate the anthropogenically forced climate change, partly determined by temporary exceedences over the Paris
Agreement aim of “well below 2°C above preindustrial levels” (Smith et al., 2018). Skillful predictions of such climate variations would therefore benefit society and substantial efforts have been devoted. For example, it has been shown that skillful prediction of ENSO is possible for 6 months to 1 year ahead (Barnston et al., 2019; Imada et al., 2015) and 2 years ahead for a particular event (Luo et al., 2008). In addition, although ENSO is one of the most important predictability sources, it has recently been demonstrated that the NAO can be predicted more than a year ahead by a dynamical climate prediction system (Dunstone et al., 2016). On decadal time scales, while a substantial component of the prediction skill originates from capturing the warming trends and temperature changes due to external forcing such as greenhouse gases and major volcanic eruptions, initialization of climate models contributes to improvement of skill in predicting sea surface temperature (SST) variations in, for example, the North Atlantic (Robson et al., 2012; Yeager et al., 2012), the North Pacific (Chikamoto et al., 2013; Mochizuki et al., 2010), and the tropical Pacific (Chikamoto et al., 2012; Meehl et al., 2016; Meehl & Teng, 2014). However, there are still challenges to achieve skillful and reliable seasonal to decadal predictions, which include improving initialization of a climate model and reducing model biases in both basic state and internal variability (Merryfield et al., 2020).

Recently, a Japanese modeling community has cooperatively developed a new ocean–atmosphere coupled model called the sixth version of Model for Interdisciplinary Research on Climate (MIROC6; Tatebe et al., 2019). There are two major updates from its predecessor MIROC5 (Watanabe et al., 2010), an official model for the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012): a finer atmospheric vertical resolution with a higher model top and incorporation of a shallow convective parameterization. Overall reproducibility of mean climate and internal climate variability on intraseasonal to decadal timescales in MIROC6 is improved over MIROC5. As for internal climate variations, intraseasonal eastward propagating Madden-Julian Oscillation (Madden & Julian, 1972) and interannual variations of the summertime East Asian monsoon are more realistically simulated in MIROC6. These improvements can be attributed to the newly implemented parameterization for shallow convection (Hirota, Ogura, et al., 2018). The quasi-biennial oscillations (QBOs) in the tropics and sudden stratospheric warming (SSW) and associated troposphere-stratosphere interactions in the Northern Hemisphere midlatitudes are also better represented, which can be attributed mainly to the higher vertical resolution in the stratosphere and the model top.

On decadal time scales, the linkage between the tropical Pacific and the Aleutian low via atmospheric teleconnections in MIROC6, which is a part of the PDV, is enhanced and more consistent with observations than in MIROC5, while reproducibility of the Atlantic multidecadal variability (AMV; Schlesinger & Ramankutty, 1994) is almost the same as in MIROC5. Regarding the mean state, biases in the tropical climate systems (e.g., underestimation of the summertime precipitation in the western Pacific and dry bias of the free troposphere) and the midlatitude atmospheric circulations both in the troposphere and the stratosphere are also significantly alleviated compared to those in MIROC5. The reader is referred to Tatebe et al. (2019) for more detailed evaluation of MIROC6 (see also Kataoka et al., 2019, for the tropical Atlantic).

The improvement listed above has motivated us to perform decadal predictions using the latest version of our coupled climate model, MIROC6. The new prediction system includes direct initialization of the atmosphere and sea ice in contrast to the MIROC5 prediction system. (MIROC5 seasonal predictions include the atmospheric initialization, but MIROC5 decadal predictions do not.) In this study, we address whether a climate model with a better representation of climate and/or initial conditions constrained by more observations lead to the improvement in seasonal to decadal prediction skill.

The paper is organized as follows. Section 2 describes the experimental design and prediction systems including climate drift. The skill of MIROC6 seasonal to interannual and decadal predictions is assessed and compared with its predecessor MIROC5 for the CMIP5 in section 3, before drawing conclusions in section 4. The decadal predictions with MIROC6 described in the paper contribute to the Decadal Climate Prediction Project (Boer et al., 2016) for the CMIP Phase 6 (CMIP6; Eyring et al., 2016) and the World Climate Research Programme Grand Challenge on Near-Term Climate Prediction (Kushnir et al., 2019).

2. Prediction System and Data

2.1. Coupled Climate Models

The state-of-the-art climate model primarily used for seasonal to decadal climate predictions is MIROC6, which has two major updates from MIROC5 as mentioned in the introduction. The model top in MIROC6...
MIROC6 is an incorporation of a shallow convective parameterization, which is a mass flux scheme based on a buoyancy-sorting, entrainment-detrainment, single-plume model of Park and Bretherton (2009). This parameterization is introduced in addition to the already implemented cumulus parameterization of Chikira and Sugiyama (2010), which deals with multiple cloud types including shallow and deep convective clouds. Shallow convection associated with atmospheric instability is calculated by the Chikira and Sugiyama (2010) scheme, and that associated with turbulence in the planetary boundary layer is represented by the Park and Bretherton (2009) scheme.

MIROC6 has several updates in its subcomponents as well. These include a nonorographic gravity wave parameterization (Hines, 1997; Watanabe et al., 2011); treatment of secondary organic aerosol and aerosols of oceanic origin (Sudo et al., 2002); an extension of the mode radius upper limit of cloud properties in the atmospheric component; physically based parameterizations for subgrid snow distribution and a snow-fed wetland that takes into consideration subgrid terrain complexity in the land surface component (Nitta et al., 2014, 2017); replacement of a warped bipolar horizontal coordinate system in the ocean component by the tripolar horizontal coordinate system with a resolution of nominal 1°, which is higher than that in MIROC5 (1.4°); enhanced oceanic vertical resolution from 49 levels to 62 levels; and improved near-surface turbulent mixing under sea ice in the oceanic component (Komuro, 2014; Murray, 1996).

2.2. Initialization

The oceanic state of MIROC6 is initialized by assimilating the global objective analysis of monthly temperature and salinity of Ishii et al. (2006) and Ishii and Kimoto (2009) in order to synchronize internal climate variations, in particular slowly varying oceanic signals which act as memories for climate predictions, in the model with those in the observations (e.g., Collins et al., 2006; Griffies & Bryan, 1998; Keenlyside et al., 2008; Mochizuki et al., 2010; Pohlmann et al., 2004). Initialization procedure for the oceanic component of MIROC6 is the same as that used in Tatebe et al. (2012) for CMIP5 decadal climate predictions with MIROC5 (Chikamoto et al., 2013; Mochizuki et al., 2012) and is performed in coupled mode. Simplified incremental analysis update (IAU; Bloom et al., 1996; Huang et al., 2002) scheme is used for initializing the upper 3,000 m of the ocean. Only anomalies of ocean temperature and salinity with respect to 1961–2000 mean are assimilated in order to minimize drifts during the predictions. Avoiding a distorted structure of the mean Atlantic meridional overturning circulation (AMOC) simulated in the full-field assimilation experiment with ocean temperature and salinity is one of the reasons for adopting so-called anomaly assimilation scheme (not shown; e.g., Smith et al., 2013).

While any observed sea ice information is not assimilated into MIROC5 for the CMIP5, sea ice concentrations in MIROC6 predictions are initialized by assimilating the absolute values of observed monthly sea ice concentration from Ishii et al. (2006) and Ishii and Kimoto (2009), which is based on satellite data (Armstrong et al., 2012), following the same procedure for ocean temperature and salinity. Since the interannual variation of the sea ice area in the Arctic region during the late fall to early winter is hypothesized to influence the occurrence of severe cold winters over Eurasia and the Far East Asia by changing paths of synoptic disturbances (e.g., Inoue et al., 2012; Mori et al., 2014, 2019; Yang et al., 2017; however, a different view is presented, e.g., in Ogawa et al., 2018, and Blackport et al., 2019), it could be a source of predictability on seasonal timescales.

Another update of the initialization for MIROC6, which was not included in MIROC5 decadal predictions, is the initialization of atmospheric zonal and meridional winds, temperature, and specific humidity. This is because stratospheric variations associated with, for example, the SSW, the QBO, and stratosphere-troposphere interactions could be key for improving seasonal forecast skill (e.g., Ineson & Scaife, 2009; Scaife, Athanassiadou, et al., 2014; Stockdale et al., 2015; Yoo & Son, 2016) but are not adequately captured by the ocean assimilation alone. As noted by Tatebe et al. (2019), there is a significant cold temperature bias in the lower troposphere over the Arctic region in MIROC6. This cold temperature bias appears to support its realistic simulation of the Arctic sea ice distribution. In a trial experiment in which the atmospheric variables are relaxed toward (full fields of) the reanalysis data during integration in coupled mode, the summer time Arctic sea ice area is significantly underestimated and the decreasing trend of the
sea ice area after 1980s is faster than the observations. Therefore, the atmospheric variables are not assimilated in the present experiments. Instead, atmospheric initial states are simply taken from the Japanese 55-year Reanalysis (JRA55; Kobayashi et al., 2015) instead of coupled assimilation runs in order to incorporate the observed atmospheric information. The JRA55 reanalysis from the surface to the 1 hPa pressure level is used except for specific humidity, which is only available up to the 100 hPa pressure level. Data from the JRA55 above these levels are extrapolated and averaged with outputs from the assimilation experiments with the model's weight linearly increasing so that the values at the top level is those from the assimilation experiments.

### 2.3. Experimental Design

Assimilation and prediction experiments for MIROC6 are conducted as follows. First, the model is integrated from 1 January 1950 to 31 December 2018 (Mochizuki et al., 2019a) with an initial condition at 1 January 1950 taken from the CMIP6 historical simulations (Tatebe & Watanabe, 2018), while being constrained by the ocean anomaly and sea ice full-field assimilations. The assimilation experiments consist of 10 ensemble members whose initial states come from different historical experiments. Then, atmospheric initial conditions for predictions with 10 ensemble members starting from 1 November every year from 1960 to 2018, as designated in Boer et al. (2016), are prepared by replacing zonal and meridional winds, temperature, and specific humidity in atmospheric initial conditions from each assimilation run with those of the JRA55 reanalysis at the same dates and times. Oceanic and land surface initial conditions for predictions are taken from each assimilation experiment. Prediction experiments are integrated for 122 months (Mochizuki et al., 2019b). Both the initialization and prediction experiments are forced with external boundary conditions that are the same as those used in the CMIP6 historical and the SSP2-4.5 scenario (Shiogama et al., 2019) simulations. Also, it is noted that a large ensemble (50 members) of the CMIP6 historical and SSP2-4.5 scenario simulations with MIROC6 is utilized for assessing the impact of initialization. In this study, the historical and scenario experiments together are referred to simply as the historical experiment. The experimental design is summarized in Table 1.

![Table 1: Summary of Experimental Design](image)

|                      | MIROC6 hindcast | MIROC5-SP | MIROC5-DP | MIROC6 historical |
|----------------------|-----------------|-----------|-----------|-------------------|
| Ocean Initialization | T, S (anomaly)  | T, S (anomaly) | T, S (anomaly) | -                |
| Sea Ice Initialization| SIC (full)      | -         | -         | -                 |
| Atmosphere Initialization| U, V, T, q (full; JRA55) | U, V, T, q (full; NCEP/NCAR) | U, V, T, q (full; JRA55) | -                |
| Initialization Date  | 1 November      | 1 November | 1 January | -                 |
| Ensemble Size        | 10 (40 for 2008–10) | 8         | 6         | 50                |

Figure 1 presents the time series of the global-mean surface air temperature (SAT) anomaly from HadCUT4 and the ensemble means of the MIROC6 assimilation and uninitialized historical experiments. Both the assimilation and the historical experiments capture the global warming trend with nearly 1 K increases during the period. These experiments are able to reproduce the temporal cooling associated with major volcanic eruptions as well, although the historical experiments tend to overestimate the global cooling in response to Mt. Pinatubo in the 1990s. While the initialization experiments assimilate only oceanic fields, it also captures the observed interannual-to-decadal fluctuations, indicating that the ocean assimilation is performing as expected including the land area.

To check the assimilation performance in the sea ice variability, the time series of the Arctic sea ice area anomaly in September, when the Arctic sea ice amount is smallest, are computed (Figure 2). It is found that the ensemble mean of the historical experiment replicates a declining long-term trend. Although the magnitude of the variation is modest compared to the observations, the assimilation experiment is able to capture the observed interannual variations as well, with an anomaly
Seasonal to decadal prediction skill of MIROC6 are evaluated in the following section and are compared with those of MIROC5 predictions, whose configurations for initialization are different from MIROC6 predictions. Prediction experiments with MIROC5 used for comparison of seasonal prediction skill is compiled by Imada et al. (2015) for the Climate-system Historical Forecast Project (CHFP; e.g., Butler et al., 2016; Tompkins et al., 2017). Hereafter, this seasonal prediction experiment is referred to as MIROC5-SP. We use MIROC5-SP, instead of the CMIP5 decadal prediction experiments with MIROC5 (hereafter, MIROC5-DP), for seasonal prediction skill comparison because of its closer configuration to MIROC6 hindcasts: the same initialization date, a closer ensemble size, and a closer atmospheric initialization. Initialization procedure for MIROC5-SP is the same as that used for MIROC6 predictions except that there is no sea ice data assimilation and that initial states of zonal and meridional winds, temperature and specific humidity are taken from the National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis (Kalnay et al., 1996). MIROC5-SP is composed of 8 member ensembles initialized each 1 November from 1980 to 2015 and integrated for 1 year.

Meanwhile, MIROC5-DP is used for comparison of decadal prediction skill. As noted earlier, neither sea ice data assimilation nor atmospheric initialization is performed for the MIROC5-DP (Tatebe et al., 2012). The MIROC5-DP, initialized every 1 January from 1960 to 2018, is integrated for 10 years with six-member ensembles. The CMIP5 historical climate and RCP4.5 scenario forcing data are used for all the MIROC5 experiments.

2.4. Evaluating Prediction Skill

It is known that climate models suffer from initialization shock and drift after initializing forecasts because of systematic errors inherent in the model (e.g., Sanchez-Gomez et al., 2016). Although anomaly assimilation is adopted for ocean temperature and salinity in our prediction system, drifts are inevitable since assimilated anomalies from observations, including anthropogenically forced trends, are not necessarily within the model attractor (e.g., Chikamoto et al., 2019). In addition, a full-field assimilation scheme is adopted for sea ice concentration and atmospheric initial states are taken from the JRA55 reanalysis. Figure 3 shows the drift in MIROC6 hindcasts, which is defined as the difference in the lead time-dependent climatology between ensemble means of the hindcasts and assimilation experiments. In the forecast Year 1, relatively large drifts in surface temperature are seen over the Eurasian and North American continents likely due to the replacement of the initial atmospheric state with the JRA55 reanalysis. The maximum amplitudes of the warming over the Eurasia and the cooling over the northern North America are close to 1 K. However, the most striking drift occurs in the tropical Pacific SST. The equatorial Pacific SSTs drift to a slightly cooler state for half a year after starting the forecast, after which El Niño begins to develop (Figure 3d). The El Niño matures at the end of the forecast Year 2 and remains until the beginning of the Year 4. As a result, the maximum systematic error (drift) in annual mean SST is seen in the forecast Year 3 with its peak as high as 1.7 K, which exceeds the observed standard deviation of the Niño3.4 index. The influence of this model drift in the equatorial Pacific is not confined to the tropics but extends into the extratropics through atmospheric teleconnections (e.g., Alexander et al., 2002). For example, in the forecast Year 3, negative sea level pressure (SLP) anomalies associated with the Pacific-North American pattern (Wallace & Gutzler, 1981) are seen over the North Pacific. The anomalous southeasterlies associated with this negative SLP anomalies cause SAT warming over northwest America, western Canada, and Alaska. Although the systematic El Niño-like drift begins to decay in the forecast Year 4 and becomes rather stable afterward, there are still fluctuations that can be comparable to decadal climate signals of interest. Therefore, there is a need to remove such systematic errors to accurately evaluate the hindcast skill. Following Boer et al. (2016), this is achieved by transforming raw outputs into anomalies with respect to the lead time-dependent forecast.
climatology:  

\[ \overline{Y}_{kj} = \overline{Y}_{kj} - \overline{Y} \]

where \( \overline{Y}_{kj} \) represents an ensemble member \( k \) of the forecast from start year \( j \) at a lead time \( \tau \) and \( \overline{Y} \) denotes averages over start years and ensembles for a given \( \tau \). Observed anomalies, against which the forecasts are to be compared, are similarly defined relative to the climatology whose time span exactly matches that used for the forecast climatology.

The ACC and the RMSE skill scores are used to evaluate skill. The impact of initialization is evaluated using “partial skill” proposed by Smith et al. (2019). The observed (\( \overline{O'} \)) and ensemble mean initialized forecast anomalies (\( \overline{Y'} \)) are decomposed as

\[ \overline{O'} = \overline{O'}_{\text{res}} Y' = \overline{Y'} + Y'_{\text{res}} \]

where \( \overline{O'} \) and \( \overline{Y'} \) are the components that can be explained by linear regression with the ensemble mean of the uninitialized experiments and the residuals \( O'_{\text{res}} \) and \( Y'_{\text{res}} \) are the variability that cannot be captured by the uninitialized experiments. The ACC and the RMSE between observed and forecast residuals are computed to measure the impact of initialization and are referred to as the partial ACC and the partial RMSE, respectively. In this study, the ACC and the RMSE between observed and forecast total anomalies are called the total ACC and the total RMSE or simply the ACC and the RMSE. Since the partial skill is found to sensitive to ensemble size of uninitialized experiments and a large ensemble is required to obtain the correct partial skill (not shown), the impact of initialization is not assessed for the MIROC5 prediction systems, in which the historical experiment is composed of only five member ensembles.
Statistical significance is evaluated using a nonparametric block bootstrap approach (Goddard et al., 2013; Wilks, 2011) with a block length of five consecutive cases, in which we obtain a probability distribution of each score based on 1,000 repetitions. The total and partial ACC and their difference between the two systems are tested whether they are larger or smaller than 0, the total and partial RMSE are tested whether they are larger or smaller than the standard deviation of the observed anomaly and residual, respectively, and the RMSE ratio is tested whether they are larger or smaller than 1. Unless otherwise noted, significance is assessed at the 10% level by a one-sided test. Also, we compute the ACC and RMSE values at each grid point for the ensemble mean of randomly chosen 10 historical simulations without replacement. The resampling is performed 10,000 times to generate a probability distribution, and skill scores with the uninitialized experiment are estimated as their median.

The observations used for hindcast evaluation are ocean temperature and sea ice concentration from Ishii et al. (2006) and Ishii and Kimoto (2009); the JRA55 reanalysis for SAT (only for seasonal to interannual predictions), zonal wind, and SLP; HadCRUT4 (Morice et al., 2012) version 4.6.0; and precipitation from the Global Precipitation Climatology Project (GPCP; Adler et al., 2003) version 2.3 for seasonal to interannual predictions and the Global Precipitation Climatology Center (GPCC; Schneider et al., 2011) full data reanalysis version 2018 for decadal predictions.

3. An Evaluation of MIROC6 Predictions

In this section, an evaluation of the MIROC6 predictions is presented. Section 3.1 focuses on the seasonal to interannual predictions, whereas section 3.2 provides an evaluation of the decadal predictions. Possible impacts of a large ensemble on seasonal forecasts are also discussed in section 3.1. Since data sets after the satellite era are expected to be more reliable and the MIROC5-SP are available only after 1979, we focus on hindcasts issued in and after 1980, whereas the MIROC6 and MIROC5-SP hindcasts initialized between 1960 and 2008 inclusive are used for the evaluation of decadal predictions. All fields presented in the decadal predictions are annual averages.

3.1. Seasonal to Interannual Predictions

3.1.1. ENSO, SAT, and SLP

We begin by examining the predictability of the dominant interannual climate variability ENSO, which influences the global climate (e.g., Alexander et al., 2002). Figure 4a shows the time series of Niño3.4 index (SST anomalies averaged over the region 170°W–120°W, 5°S–5°N) from the observations and MIROC hindcasts. It is found that both MIROC hindcasts are able to capture the overall phase of the observed ENSO, but MIROC6 outperforms MIROC5-SP for forecasting the development and/or decay of major ENSO events such as 1982/83 El Niño decay, 1997/98 El Niño development and decay, 2010/11 La Niña decay, and 2014/2015 El Niño growth. As a result, significant ACC lasts longer and RMSE is significantly reduced for the lead times up to 2 months according to the RMSE ratio test (Figures 4b and 4c). Essentially the same RMSE at longer lead times despite the ACC from the MIROC6 hindcasts being better for lead times longer than 6 months is because the MIROC6 hindcasts tend to overestimate the Niño3.4 amplitude at longer leads. MIROC6, however, fails to track the observed time series for some cases (e.g., hindcasts starting in 1983 and 2009) and skill scores themselves are modest compared to the hindcasts with other state-of-the-art models (e.g., Barnston et al., 2019). It should be noted that the drop in skill in Spring (5 months lead time corresponds to April) is common to other prediction systems and is known as the spring predictability barrier (e.g., Jin et al., 2008; Wang et al., 2019).

Skill maps for hindcasts of the first boreal winter (December–February; DJF) and summer (June–August) SAT and SLP are shown in Figures 5 and 6, respectively. (Hereafter, seasons refer to those for the Northern Hemisphere.) Significant prediction skill is seen for the MIROC6 hindcasts of the first winter SAT and SLP almost all over the world and they tend to be improved relative to MIROC5-SP, particularly over the tropical oceans. ACC (RMSE) scores of the MIROC6 hindcast are improved over the MIROC5-SP in about 58% (57%) area of the globe for SAT and about 67% (67%) area for SLP. One of the notable improvements over the MIROC5-SP is enhanced skill scores for the SLP over the North Atlantic. The MIROC6 hindcasts show positive ACCs around the nodal regions of the NAO, although the ACCs over its northern pole are not statistically significant. As a result, although the value itself is modest, MIROC6 predictions show larger ACC skill score for the first winter NAO: ACC scores of MIROC6 and MIROC5-SP are 0.3 (significant at the 10% level.
[here block length for the significance test is set to 1 case as the NAO does not show a significant autocorrelation; Dunstone et al., 2016]) and 0.14, respectively. The improvement is perhaps a reflection of a higher atmospheric vertical resolution and model top in MIROC6, but a better representation of the ENSO teleconnection may also contribute to it as will be discussed later. Recent studies suggest that climate models tend to underestimate the predictable NAO signals relative to atmospheric internal “noise” and large ensembles are required for skillful NAO predictions (so-called “signal-to-noise paradox”; Dunstone et al., 2016; Eade et al., 2014; Scaife & Smith, 2018). To test this possibility, the ratio of predictable components (RPCs; Eade et al., 2014) between the observations and MIROC6 hindcasts is calculated. Here RPC is defined as the ratio of the correlation between the model ensemble mean and the observations to the average correlation between the model ensemble mean and a single ensemble member (Scaife & Smith, 2018). The RPC value for the NAO in MIROC6 hindcasts is 2.9. This RPC value greater than 1 implies that MIROC6 hindcasts underestimate the predictable signal and/or overestimate the unpredictable noise. Thus, the signal-to-noise paradox applies to the NAO in MIROC6 hindcasts and relatively small size ensembles (10 members) may explain the significant but relatively low ACC in MIROC6 NAO hindcasts. To give a further insight into this issue, we have prepared additional 30 ensembles for three selected cases (hindcasts issued from November 2008 to 2010) integrated through 2 years. Figure 7 shows the pattern correlations (the uncentered correlation is used to take signs into account [DelSole & Shukla, 2006], but conclusions are unchanged even if the centered pattern correlation is used; also, the result is not sensitive to a slight shift in the domain size) of SLP anomalies over the North Atlantic between observations and the ensemble means of the total and the original hindcasts. The
correlation coefficients for the 40-member mean are significantly improved over the original 10-member mean for both the first and second winters of all three cases with the only exception of the second winter of 2010/2011. In particular, while the pattern correlation for the second winter of 2009/2010 with the original 10 ensembles is the worst among the MIROC6 hindcasts initialized after 1979 at less than −0.8, the correlation becomes as high as 0.9 when the ensemble is increased to 40. Thus, we may conclude that the increased ensembles appear to enhance the NAO predictability for both the first and second winters as for other prediction systems (e.g., Dunstone et al., 2016; Scaife, Arribas, et al., 2014).

Figure 5. ACC of seasonal-mean SAT from the MIROC6 hindcasts for (a) the first boreal winter (December–February) and (b) the first boreal summer (June–August). (c, d) As in (a) and (b) but for ACC differences between the MIROC6 hindcast and the MIROC5-SP. (e, f) As in (a) and (b) but for RMSE ratio of the MIROC5-SP to the MIROC6 hindcast: Values >1 indicate the RMSE is reduced in MIROC6. Stippling denotes regions where the values are significant at the 10% level. In (c)–(f), the fraction of areas where the skill with MIROC6 is improved over MIROC5-SP is indicated at the top of each panel.
The western equatorial Pacific SAT RMSEs are reduced in MIROC6 hindcasts (Figure 5). Regression analysis of winter SAT anomalies onto the DJF-mean Niño3.4 index reveals that it is a reflection of an improvement in representation of the spatial structure of ENSO (Figure S1 in the supporting information). The centers of warm anomalies along the equator are between approximately 160°W and 110°W both in the observations and MIROC6 hindcasts. Anomalies in MIROC5-SP, however, are zonally elongated and relatively large anomalies extend beyond the dateline, which is a common bias of coupled climate models (e.g., Guilyardi et al., 2009), resulting in ENSO SST prediction errors in the western equatorial Pacific. Regarding the improvements in SLP, the same analysis but for SLP anomalies suggests a correct representation of the ENSO-NAO relationship may contribute to a larger ACC for the MIROC6 hindcasts of the NAO; the negative NAO tends to co-occur with the El Niño in the observations and the MIROC6 hindcasts but the positive NAO

Figure 6. As in Figure 5 but for SLP.
Figure 7. Pattern correlations of winter SLP anomalies over the North Atlantic (60°W–0°, 30–80°N) between observations and the ensemble means of the original 10-member ensembles (blue lines) and 40-member ensembles (red lines) for three selected cases where we have large means of the original 10-member skill at the 10% level are indicated by open circles.

Regarding the first summer predictions, an improvement of MIROC6 hindcasts over MIROC5-SP is not clear unlike that for the first winter prediction skill. For example, while the western tropical Pacific SAT RMSEs/ACCs are smaller/larger in MIROC6, which again is associated with the better representation of ENSO pattern (Figures S1j, S1m, and S1p), those in the eastern part are slightly worse. Associated with this, the MIROC6 shows better ACC skill for precipitation in the western tropical Pacific but worse in the eastern part. Also, although MIROC6 shows a better skill in summer SLP prediction over the Pacific, SLP over the Atlantic is generally better predicted in MIROC5-SP. In addition to the tropical regions, significant ACCs for SAT are found over Greenland, China, and from eastern Europe to Arabian Peninsula, but they are originated mostly in capturing the global warming trend (not shown).

3.1.2. Arctic Sea Ice

One of the updates of our prediction system is initialization of the sea ice, which is not considered in MIROC5-SP. Spatial maps of skill scores for the Arctic sea ice concentration from the MIROC6 hindcasts are presented in Figure 8. Significant ACCs for the first winter are found over the Hudson Bay, the Labrador Sea, the Iceland Sea, the Kara Seas, and the Sea of Okhotsk, among which skill over the Hudson Bay, the Barents Sea, and the Sea of Okhotsk is enhanced compared to the MIROC5-SP, and 57% of the Arctic sea ice region shows the skill improvement associated with the system update (Figures 8a and 8c). We note that variability and trends in winter sea ice concentration are very small across most of the central Arctic, where concentrations are close to 100% in winter, and therefore, places with small standard deviations of concentrations are masked out in order to focus on regions of large variability (and thus may affect the atmosphere). Also, the RMSE values are improved in the Hudson Bay, the northwestern Labrador Sea, the Iceland Sea, the Barents Sea, and the Sea of Okhotsk and area with skill improvement from the MIROC5-SP to the MIROC6 is 65% (Figure 8e). While the MIROC6 hindcast skill arises from capturing the sea ice response to the external forcing (i.e., decreasing trends) to a certain extent, it is improved by initialization over the regions including the Hudson Bay, the Labrador Sea, the Iceland Sea, the Barents Sea, and the Sea of Okhotsk for the ACC and the Labrador Sea, the Barents Sea, and the Sea of Okhotsk for the RMSE (Figure S3). While previous studies suggest that sea ice cover variations over Barents-Kara Seas around autumn and winter may influence the climate over the Eurasia during winter (Inoue et al., 2012; Mori et al., 2014), the significant prediction skill in MIROC6 does not lead to significant ACCs or clear improvements over MIROC5-SP for winter SAT and SLP over Eurasia (Figures 5 and 6). We note that this is an active research area, and larger ensembles may be required to obtain robust results (Mori et al., 2019). At 10 months lead time (September), RMSE values in MIROC6 hindcasts over the Barents-Kara Seas are reduced compared to MIROC5-SP because MIROC5-SP produces spurious sea ice variability, where there is no sea ice variation in the observations (Figures 8f). Although statistically significant ACCs are found
Figure 8. (a, b) As in Figure 5a but for sea ice concentration anomalies from the MIROC6 hindcasts for the first winter and September. Regions with the standard deviation < 0.03 in both the observations and predictions are not drawn. (c, d) As in (a)–(c) but for ACC differences between the MIROC6 hindcast and the MIROC5-SP. (e, f) As in (a) and (b) but for RMSE ratio of the MIROC5-SP to MIROC6 hindcast. Regions where RMSE values from both the MIROC5-SP and MIROC6 hindcasts are zero are not drawn.
over a large domain, they arise mostly from capturing externally forced trends (Figures 8a and S3). However, MIRCO6 shows better ACC around the North Pole with significant impact of initialization (Figures 8d and S3), and the more than 60% area shows better skill compared to the MIROC5-SP for both the ACC and the RMSE (Figures 8d and 8f).

3.1.3. Quasi-Biennial Oscillation
Since the QBO has predictability beyond a year (Scaife, Athanassiadou, et al., 2014), it is interesting to investigate if our prediction systems can skillfully predict its behavior. In the JRA55 reanalysis, zonally averaged equatorial zonal wind anomalies show downward propagating easterly and westerly regimes oscillating with an average period of about 28 months (Figure 9a; Baldwin et al., 2001). The same time series but from the MIROC6 hindcasts in the forecast Year 1 shows that the MIROC6 hindcasts are capable of capturing the overall QBO phase including the downward propagating feature, though the amplitudes are modest at 10–30 hPa and diminished in the lowermost stratosphere compared to the JRA55 data (Figure 9b). We note that the

Figure 9. Time-height cross sections of the monthly zonal-mean zonal wind anomalies over the equator (5°S–5°N average) from (a) JRA55, (b) MIROC6 hindcasts, and (c) MIROC5-SP during the forecast Year 1. Contour intervals are 3 m/s, and gray vertical line denotes November 1st in each year. (d) Anomaly correlations and (e) RMSE of zonal-mean zonal wind anomalies over the equator from the MIROC6 hindcasts as a function of lead time. (f–i) As in (d) and (e) but for (f, g) the MIROC5-SP and (h, i) the MIROC6 historical experiment. Stippling denotes regions where skill scores are significant at the 10% level.
MIROC6 does not capture the QBO disruption in 2016 (Newman et al., 2016; Osprey et al., 2016). A higher vertical resolution in the stratosphere (Watanabe et al., 2018) and/or a more accurate prediction of the Barents-Kara sea ice variability (Hirota, Shiogama, et al., 2018) may be required for a successful prediction of this event. On the other hand, MIROC5-SP predictions do not reproduce any oscillation or downward propagation of the initial anomalies (Figure 9c). This is expected because while the QBO is absent in MIROC5 (Butchart et al., 2018), MIROC6 is able to simulate a realistic QBO owing to its better stratospheric resolution and nonorographic gravity wave drag parameterization (Tatebe et al., 2019). As a result, the ACC/RMSE (Butchart et al., 2018), MIROC6 is able to simulate a realistic QBO owing to its better stratospheric resolution and nonorographic gravity wave drag parameterization (Tatebe et al., 2019). Surprisingly, the ensemble mean of the uninitialized historical experiments also shows a predictive skill with a maximum ACC more than 0.7 despite no clear trends in neither JRA55 nor MIROC6 historical experiments. A large ensemble (50 members) average of MIROC6 historical experiments actually displays a QBO-like cycle in equatorial zonal winds (Figure S4). What drives the QBO in MIROC6 historical experiment? To answer this question, three additional historical experiments are conducted following the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al., 2016) protocol: hist-aer (Shiogama, 2019a), hist-GHG (Shiogama, 2019b), and hist-stratO3 (Shiogama, 2019c) experiments. The hist-stratO3 experiment, in which MIROC6 is forced only with the historical stratospheric ozone data, reproduces the historical QBO (Figure S5). Since the CMIP6 ozone data (at least partially) includes ozone QBO signal (Pohlmann et al., 2019), it appears that diabatic heating anomalies associated with prescribed stratospheric ozone anomalies would feed back onto the dynamical QBO (e.g., Naoe et al., 2017). The details how the ozone QBO drives the dynamical QBO will be reported elsewhere. As the “future” QBO signal is of course what we want to predict, the climatological or fixed ozone concentration should be specified to assess actual QBO forecast skill (or interactive ozone chemistry should be included). Nonetheless, since ACCs and RMSEs in MIROC6 hindcasts are improved over the historical experiment up to a lead time of 3 years, we may conclude that there is a relatively long-term QBO predictability in MIROC6 predictions. However, our hindcasts fail to reproduce the extratropical response to the QBO in winter (Boer & Hamilton, 2008; Ebdon, 1975; Holton & Tan, 1980; Scaife, Athanassiadou, et al., 2014), so that the skillful prediction of the QBO, though partially owing to the diabatic heating effect of prescribed ozone, does not contribute to an additional skill for surface winter climate predictions at least in our system.

3.2. Decadal Predictions
3.2.1. Global Surface Temperature

Time series of globally averaged annual mean SAT from the observation and MIROC6 hindcasts at lead times of Years 2–5 and 6–9 are presented along with the MIROC6 historical experiments in Figure 10a. MIROC6 hindcasts are able to capture the global warming trend and show substantial skill in predicting the global mean SAT (ACC > 0.9; Figure 10b). Although MIROC6 hindcasts for 2–5 years lead time overestimate the warming rate in the early 21st century, so-called global warming hiatus period (Easterling & Wehner, 2009), mainly due to excessive warming over North America and Eurasia, they better predict the cooling response to Mt. Pinatubo eruption around 1990 than the uninitialized experiment. The total RMSE at this lead time is significantly reduced compared to the MIROC5-DP and the significant impact of initialization (partial ACC) is seen until a lead time of 3–6 years (Figures 10b and 10c). For 6–9 years lead time, MIROC6 hindcasts are generally close to the ensemble mean of the historical experiment.

Consistent with the substantial global mean SAT skill, significant surface temperature skill is found almost everywhere with exceptions such as the northeast Pacific and the Southern Ocean (Figures 11a and 11b). While total ACCs for global mean SAT are more or less similar across the whole lead times between the two systems, the MIROC6 hindcasts have a larger fraction of areas with better skill both for total ACC and RMSE at all lead times (Figures 11c–11f). Regarding the impact of initialization, temperature skill is improved over the North Pacific for 2–5 years lead time and the eastern Indian Ocean and the North Atlantic Ocean for both 2–5 and 6–9 years lead times (Figures 12a and 12b). In the following, hindcast skill for the North Atlantic and Pacific Oceans is investigated in more detail.
3.2.2. Atlantic Multidecadal Variability

The North Atlantic Ocean exhibits decadal variations called AMV and is known to be a region where climate models are capable of predicting its variability on multiyear to decadal timescales (Chikamoto et al., 2013; Doblas-Reyes et al., 2013; Keenlyside et al., 2008; Mochizuki et al., 2012; Robson, Hodson, et al., 2014; Shaffrey et al., 2018; Yeager et al., 2012). Since the North Atlantic Ocean variability on these timescales appears to be a combination of internal dynamics and external forcing (e.g., Booth et al., 2012; Robson et al., 2012, 2016; Watanabe & Tatebe, 2019; Yeager et al., 2012), both initialization and external forcing are important for its prediction. Observed AMV index (Trenberth & Shea, 2006; SST anomalies average over 0°–60°N, 80°W–0° minus global mean SST anomaly) undergoes cool conditions during the 1970s to the 80s and switches to a warm condition after mid-1990s (Figure 10d). This behavior is well captured by MIROC6 hindcasts at 2–5 years lead time and to a lesser extent at 6–9 years lead time. Although the MIROC6 hindcasts produce skillful predictions of the AMV index across lead times up to 7–10 years with the ACC more than 0.6, the skill scores from the MIROC5-DP are generally better than the MIROC6 hindcasts (Figures 10e and 10f; difference is significant for the ACC at 5–8 years lead time and for the RMSE at 5–8 and 6–9 years lead times). Also, while the MIROC6 hindcasts are able to capture the overall fluctuations, the ensemble mean of the uninitialized MIROC6 historical experiment also reproduces the observed AMV index variations with a high ACC of about 0.7. Thus, a large portion of the skill for the AMV index in MIROC6 hindcasts arises from capturing the externally forced component. However, the MIROC6 hindcasts for the AMV index are significantly improved by the initialization up to 3–6 lead years.

Significant skill improvements associated with initialization are seen for the SST variability over the North Atlantic subpolar gyre up to 4–7 years lead time (after which it is confined in the Labrador Sea region and the

![Figure 10.](https://example.com/figure10.png)
northern part) and over the tropical North Atlantic across the whole lead times (Figures 12a and 12b). However, we note that although the significant total and partial SLP skill is also found over the latter region (Figures 12c, 12d, 13a, and 13b), the source of the predictability is not clear. On the other hand, previous studies show that the anomalously weak and strong AMOC are important respectively in the

---

**Figure 11.** As in Figure 5 but for annual-mean SST (over oceans) and SAT (over land) anomalies at the lead time of (a, c, and e) 2–5 and (b, d, and f) 6–9 years. Note that the MIROC5-DP is used here. Stippling denotes regions where the values are significant at the 10% level. In (c)–(f), the fraction of areas where the skill with MIROC6 is improved over MIROC5-SP is indicated at the top of each panel.
rapid cooling in the 1960s and warming in the 1990s over the North Atlantic subpolar gyre region (Robson et al., 2012; Robson, Hodson, et al., 2014; Yeager et al., 2012). MIROC6 hindcasts at 2–5 years lead time reproduce the AMOC decline in the mid-1960s and stronger AMOC during the 1990s with a peak around the mid-1990s (Figure S6). In contrast, both MIROC6 hindcasts at 6–9 years lead time and the uninitialized historical experiment do not capture these AMOC variations. Thus, we speculate that the improvement of the skill for the North Atlantic subpolar gyre SSTs, where the internal climate dynamics dominates (Robson, Hodson, et al., 2014; Watanabe & Tatebe, 2019; Yeager et al., 2012), can be attributed to predictability associated with the initialization of the AMOC conditions. We note that the 1960s decline in the proxy for the AMOC from oceanic data used for the initialization of MIROC6 hindcasts is modest compared to that of the predicted AMOC volume transport at 2–5 years lead time (black and red lines in Figure S6) and the AMOC proxy used in the previous studies, which is also based on the Labrador Sea water density (Robson, Hodson, et al., 2014; Robson et al., 2016). The difference in proxies may be due to the differences between the datasets and/or depth for averaging in each study.

### 3.2.3. Pacific Decadal Variability

Climate fluctuations on decadal timescales are observed in the Pacific as well. The PDV, including both the interdecadal Pacific oscillation (Power et al., 1999) and the Pacific decadal oscillation (PDO; Mantua et al., 1997), is suggested to have multidecadal timescales predictability (Chikamoto et al., 2012, 2013; Meehl & Teng, 2012, 2014; Mochizuki et al., 2010, 2012; van Oldenborgh et al., 2012). The black line in Figure 10g shows the observed PDO index. The cooler states from the 1960s to the early 1970s shifted to warmer conditions in the late 1970s, which has switched back to a cool condition during the recent hiatus period. Although the MIROC6 hindcasts tend to underestimate the amplitude, the cooler conditions around the 2000s, the warmer state around the 1980s, and some cooling in the 1960s are reasonably captured at the 2–5 years lead time. Skill for predicting the PDO index in MIROC6 at this lead time is significantly improved by initialization (Figures 10h and 10i). In contrast, the MIROC5-DP does not show a significant skill for the PDO index at 2–5 years lead time, although the ACC at a lead time of 3–6 years is significant. The improvements by initialization in the MIROC6 hindcasts are seen not only in the North Pacific but also in the tropical Pacific.
(Figure 12a). ACCs in some locations of the North Pacific including east of Japan and RMSEs in the tropics are improved relative to the MIROC5-DP (Figures 11c and 11e). The enhanced skill in the North Pacific may be due to a better representation of the subtropical gyres including the western boundary currents in MIROC6 (Tatebe et al., 2019).

Closer inspection reveals that the predictive skill for the PDO index in MIROC6 hindcasts is related to a more realistic prediction of the climate regime shift of the Pacific in the late 1970s (Figures 14a–14d). As mentioned

Figure 13. As in Figure 11 but for annual-mean SLP anomalies.
above, the Pacific experienced a regime shift toward a positive phase of the PDV in the late 1970s with an El Niño-like warming in the tropics and subtropical cooling sandwiched by the warming in the northern and southeastern North Pacific. Although the MIROC6 hindcasts do not capture the central Pacific cooling in lower latitudes, they are able to reproduce the abovementioned features of the transition toward the positive PDV with a (uncentered) pattern correlation of 0.55, which is significant at the 10% level. The MIROC5-DP also shows subtropical cooling in the North Pacific, but its center is shifted to the east.

Figure 14. Spatial distribution of SST changes (in K) for the late-1970s regime shift represented by the composite difference between Year 2-5 hindcasts verifying during the periods 1979–1992 and 1961–1973 from (a) the MIROC6 hindcasts and (b) the MIROC5-DP. (c) As in (a) but for the verifying observations. (d) Difference between (a) and (b). Stippling denotes regions where the difference is significant at the 10% level by a two-sided test. Significance for the observations is assessed using a standard t test. The two numbers at the top right corner of the panels (a), (b), and (d) indicate the centered and uncentered pattern correlation against the observations over the domain 120°E–80°W, 30°S–70°N. Note that the composite periods for the MIROC5-DP are shifted by 2 months because of the difference in the initialization date. (c, f) As in Figures 10b and 10c but for annual mean Niño3 SST anomalies.
Moreover, tropical warming is nearly uniform. As a result, the MIROC6 hindcasts better capture the spatial structure including the cooling east of Japan, the warming off the west coast of North America, and the eastern tropical Pacific warming, with a significantly higher pattern correlation. (Both centered and uncentered correlations are significantly improved over the MIROC5-DP.) Also, the eastern tropical Pacific warming in the 50-member mean of the MIROC6 uninitialized experiments is about 60% of that in the MIROC6 hindcasts and it fails to capture the cooling in the subtropical North Pacific (not shown), highlighting the importance of initialization in the MIROC6 hindcasts.

Although the MIROC6 hindcasts reasonably reproduce the 1970s regime shift in the Pacific SSTs and the changes in the SLP over the tropics (i.e., weakening of the Walker Circulation), the associated deepening of the Aleutian low is not captured (not shown). A better representation of the atmospheric teleconnections to the extratropics may contribute to predictive skill for the North Pacific SSTs at longer lead times since the MIROC6 hindcasts show significant skill in predicting annual-mean Niño3 SST up to 3–6 years lead time (Figure 14e). Note that significant skill is also seen at longer lead times, but this may be due to sampling noise as there is no apparent physical mechanism for a return of skill improvement. The better connection of the equatorial Pacific to the extratropics may also enhance any predictability in the global mean SAT given that the tropical Pacific is a key region for the global mean SAT variability (Bordbar et al., 2019; Kosaka & Xie, 2013, 2016; Watanabe et al., 2014). Finally, we note that initialization of MIROC6 hindcasts do not contribute to an improvement in skill for a recently identified tropical Pacific-Atlantic seesaw of the SST anomalies that involves the modulation of the Walker Circulation (Chikamoto et al., 2015).

4. Conclusions and Discussion

We have developed a new seasonal-to-decadal climate prediction system based on a coupled climate model MIROC6 that contributed to CMIP6. This system adopts an initialization scheme same as a previous prediction system based on the CMIP5 version model (MIROC5-SP and MIROC5-DP), but the updates in the model (e.g., high-top atmosphere resolving the stratosphere) and the initialization of the sea ice and the atmosphere newly incorporated are expected to provide sources of extended predictability of climate variations on a wide range of the time scales. Evaluation of the MIROC6 prediction system using 10-member hindcast experiments for 1960–2018, initialized every year on 1 November, leads to the following conclusions.

Compared to a previous prediction system (MIROC5-SP), the MIROC6 prediction system shows improved skill of the ENSO prediction especially for the major events such as 1997/1998 El Niño, the NAO, the QBO, and the Arctic sea ice variability at seasonal time scale. Specifically,

1. The QBO can be predicted with the lead time of up to 3 years, with the maximum ACC exceeding \( r = 0.8 \), owing to the improved representation of the QBO in MIROC6. At the same time, the high skill for the QBO prediction partially arises from diabatic heating effect of the prescribed stratospheric ozone concentration anomalies that act to influence the dynamical QBO.

2. A statistically significant skill is obtained for the first winter of the NAO, although the ACC value itself is modest. When the ensemble size is increased up to 40, the prediction skill is enhanced for both the first and second winters, consistent with previous studies. Skillful prediction of the QBO does not provide a source of the NAO predictability.

3. About 60% of the Arctic sea ice region in both the first winter and September shows the skill improvement associated with the system update. In particular, arctic sea ice concentration in the first winter is well predicted over the Hudson Bay, Barents Sea, and the Sea of Okhotsk and the skill is improved over MIROC5-SP. However, it does not lead to an improved prediction of the winter Eurasian climate anomalies that may have partially been forced by the Barents-Kara sea ice variability.

The MIROC6 decadal hindcasts were performed by extending the seasonal hindcasts, and the skill was compared to the existing MIROC5-DP for annual-mean SAT and SST anomalies associated with major modes of decadal variability such as the PDV and AMV. Initialization improves the decadal prediction skill for the global-mean SAT with the MIROC6 up to 3–6 years lead time, and the MIROC6 hindcast has a larger fraction of areas with better skill both for ACC and RMSE at all lead times compared to the MIROC5-DP. In addition, several improvements in predicting the basin-scale variability are identified.
1. SST anomalies in the North Atlantic subpolar gyre are predictable 4–7 years in advance. Successful initialization of the AMOC is likely to contribute to the high-latitude SST anomalies being predicted. There is a significant prediction skill (ACC > 0.6) for the AMV index with the lead time of several years, but the RMSE reaches the level of uninitialized experiment; it supports previous studies showing that the past AMV was driven partially by external forcing.

2. Predictability of the Pacific SSTs at decadal time scales is limited, but the skill is slightly increased in the MIROC6 decadal predictions. At a lead time of 2–5 years, MIROC6 decadal predictions better track the PDV phase reversal that occurred in the late 1970s than the MIROC5-DP. In particular, the cooling east of Japan, the warming off the west coast of North America, and the eastern tropical Pacific warming are better captured.

We have shown that the new prediction system based on MIROC6 provides significant prediction skill over the ocean on multiyear timescales. However, it is desirable to understand the physical mechanism by which such a long predictability occurs. Of particular interest is the high ACC in the tropical North Atlantic SST, which has been less focused on than the SST variability over the North Atlantic subpolar gyre region but can be an important driver of the climate system (e.g., Ham et al., 2013). This issue will be addressed in a future study.

The MIROC6 prediction also has a skill at predicting annual-mean precipitation in a few locations including the Sahel and the eastern Eurasia (Figure S7). Although initialization significantly improves precipitation skill in these regions, skill is sporadic particularly at the longer lead time. This sporadic nature may suggest that the signal-to-noise paradox applies to our decadal predictions and lager ensembles are needed to detect a predictable component in the model (Smith et al., 2019; Yeager et al., 2018) and/or that a higher horizontal resolution is required (e.g., Kimoto et al., 2005).

**Data Availability Statement**

The GPCP and GPCC data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd; MIROC5 seasonal predictions are available from CHFP Data Archive at https://www.wcrp-climate.org/wgsp-chfp/chfp-data-archive; the JRA55 reanalysis at https://jra.kishou.go.jp/JRA-55/index_en.html/jra-55; HadCRUT4 at https://crudata.uea.ac.uk/cru/data/temperature/ website. The CMIP6 experiment data used in the present study are distributed through the Earth System Grid Federation and are freely accessible. Details on ESGF are given on the CMIP Panel website (https://www.wcrp-climate.org/wgcm-cmip). The model code and data are available upon reasonable request and should contact the corresponding author.

**Acknowledgments**

We thank three anonymous reviewers for their constructive comments and suggestions that led to a substantial improvement of this manuscript. We also thank Yoshio Kawatani, Yoshiki Komuro, Masato Mori, Jun Ono, Tatsuo Suzuki, and Shingo Watanabe for their helpful discussions and comments and Masayoshi Ishii for providing the oceanic and sea ice concentration data. This work was supported by the Integrated Research Program for Advancing Climate Models (TOUGOU) Grant JPMXD171935457 from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. Model simulations were performed on the Earth Simulator at JAMSTEC.

**References**

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-F., Janowiak, J., et al. (2003). The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, 4, 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., & Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, 15(16), 2205–2221. https://doi.org/10.1175/1520-0442(2002)015<2205:TAUIAH>2.0.CO;2

Armstrong, R. L., Knowles, K. W., Brodzik, M. J., & Hardman, M. A. (2012). DMSP SSM/I FDRDaily EASE-Grid precipitation product. Boulder, Colorado, USA: National Snow and Ice Data Center, Digital Media, Earth Observing System Data and Information System (EOSDIS) Data User Services, National Snow and Ice Data Center (NSIDC), National Snow and Ice Data Center (NSIDC) snow and ice data product, (2012). Version 2.0. National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA. https://nsidc.org/data/nsidc-0032. html

Baldwin, M. P., Gray, L. J., Dunkerton, T. K., Hamilton, K., Haynes, P. H., Randel, W. J., et al. (2001). The quasi-biennial oscillation. *Reviews of Geophysics*, 39(2), 179–229. https://doi.org/10.1029/1999RG000073

Barnstom, A. G., Tippett, M. K., Ranganathan, M., & L’Heureux, M. L. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, 53, 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Blackport, R., Screen, J. A., van der Wiel, K., & Bintanja, R. (2019). Minimal influence of reduced Arctic sea ice on coincident cold winters in mid-latitudes. *Nature Climate Change*, 9(9), 697–704. https://doi.org/10.1038/s41558-019-0551-4

Bloom, S. C., Takacs, L., Da Silva, A. M., & Ledvina, D. (1996). Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124(6), 1256–1271. https://doi.org/10.1175/1520-0493(1996)124<1256:DAU>2.0.CO;2

Boer, G., & Hamilton, K. (2008). QBO influence on extratropical predictive skill. *Climate Dynamics*, 31(7–8), 987–1000. https://doi.org/10.1007/s00382-008-0379-5

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9, 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Booth, B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., & Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, 484, 228–232. https://doi.org/10.1038/nature10946
Bordbar, M. H., England, M. H., Sen Gupta, A., Santos, O., Taschetto, A. S., Martín, T., et al. (2019). Uncertainty in near-term global surface warming linked to tropical Pacific climate variability. *Nature Communications*, 10(1), 1990. https://doi.org/10.1038/s41467-019-09761-2

Butchart, N., Anstey, J. A., Hamilton, K., Ogrey, S., McLandress, C., Bushell, A. C., et al. (2018). Overview of experiment design and comparison of models participating in phase 1 of the SPARC quasi-biennial oscillation initiative (QB0). *Geoscientific Model Development*, 11, 1099–1032. https://doi.org/10.5194/gmd-11-1099-2018

Butler, A., Alves, O., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., et al. (2016). The Climate Historical Forecast Project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quarterly Journal of the Royal Meteorological Society*, 142, 1413–1427. https://doi.org/10.1002/qj.2743

Camargo, S. J., & Sobel, A. H. (2005). Western North Pacific tropical cyclone intensity and ENSO. *Journal of Climate*, 18(15), 2996–3006. https://doi.org/10.1175/JCLI3457.1

Chikamoto, Y., Kimoto, M., Ishii, M., Mochizuki, T., Sakamoto, T. T., Tatebe, H., et al. (2013). An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. *Climate Dynamics*, 40, 1201–1222. https://doi.org/10.1007/s00382-012-1351-y

Chikamoto, Y., Kimoto, M., Watanabe, M., Ishii, M., & Mochizuki, T. (2012). Relationship between the Pacific and Atlantic stepwise climate change during the 1990s. *Geophysical Research Letters*, 39, L21710. https://doi.org/10.1029/2012GL053901

Chikamoto, Y., Timmermann, A., Luo, J.-J., Mochizuki, T., Kimoto, M., Watanabe, M., et al. (2015). Skillful multi-year predictions of tropical trans-basin climate variability. *Nature Communications*, 6, 6869. https://doi.org/10.1038/ncomms7889

Chikamoto, Y., Timmermann, A., Widiansky, M. J., Zhang, S., & Balmaseda, M. A. (2019). A drift-free decadal climate prediction system for the Community Earth System Model. *Journal of Climate*, 32, 5967–5995. https://doi.org/10.1175/JCLI-D-18-0788.1

Chikira, M., & Sugiyama, M. (2010). A cumulus parameterization with state-dependent entrainment rate. Part I: Description and sensitivity to temperature and humidity profiles. *Journal of the Atmospheric Sciences*, 67, 2171–2193. https://doi.org/10.1175/2010JAS3316.1

Collins, M., Botzet, A., Carril, A. F., Hazeleger, W., Houzeau, A., Latif, M., et al. (2006). Interannual to decadal climate predictability in the North Atlantic: A multimodel ensemble study. *Journal of Climate*, 19, 1195–1203. https://doi.org/10.1175/JCLI3654.1

DeSole, T., & Shukla, J. (2006). Specification of wintertime North American surface temperature. *Journal of Climate*, 15, 2691–2716. https://doi.org/10.1175/JCLI3704.1

Delworth, T. L., Zeng, F., Vecchi, G. A., Yang, X., Zhang, L., & Zhang, R. (2016). The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience*, 9, 509–512. https://doi.org/10.1038/ngeo2738

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., Garcia-Serrano, J., Guemas, V., Kimoto, M., et al. (2013). Initialized near-term regional climate change. *Nature Communications*, 4, 1715. https://doi.org/10.1038/ncomms3204

Doi, T., Behera, S. K., & Yamagata, T. (2013). Predictability of the Niño3.4/Niña. *Scientific Reports*, 3(1). https://doi.org/10.1038/srep02882

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., et al. (2016). Skillful predictions of the winter North Atlantic oscillation one year ahead. *Nature Geoscience*, 9, 809–814. https://doi.org/10.1038/ngeo2824

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41, 5620–5628. https://doi.org/10.1002/2014GL061146

Easterling, D. R., & Wehner, M. F. (2009). Is the climate warming or cooling? *Geophysical Research Letters*, 36, L08706. https://doi.org/10.1029/2009GL037810

Edson, R. A. (1975). The quasi-biennial oscillation and its association with tropospheric circulation patterns. *Meteorological Magazine*, 104, 282–297.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Feng, M., McPhaden, M. J., Xie, S.-P., & Hafner, J. (2013). La Niña forces unprecedented Leeuwin Current warming in 2011. *Scientific Reports*, 3, 1277. https://doi.org/10.1038/srep01277

Gill, A. E. (1980). Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, 106, 447–462. https://doi.org/10.1002/qj.49710644905

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., et al. (2016). The Detection And Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geoscientific Model Development*, 9, 3685–3697. https://doi.org/10.5194/gmd-9-3685-2016

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., et al. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40, 245–272. https://doi.org/10.1007/s00382-012-1481-2

Griffies, S. M., & Bryan, K. (1998). Predictability of the North Atlantic multidecadal climate variability. *Science*, 279(5729), 181–184. https://doi.org/10.1126.science.279.5297.181

Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., et al. (2009). Understanding El Niño in ocean-atmosphere general circulation models: Progress and challenges. *Bulletin of the American Meteorological Society*, 90, 325–340. https://doi.org/10.1175/2010BAMS3309.1

Holton, J. R., & Tan, H.-C. (1980). The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, 37, 2200–2208. https://doi.org/10.1175/1520-0469(1980)037<2200:TEQBOO>2.0.CO;2

Huang, B., Kinter, J., & Sospen, P. (2002). Ocean data assimilation using intermittent analyses and continuous model error correction. *Advances in Atmospheric Sciences*, 19(6), 965–992. https://doi.org/10.1007/s00376-002-0059-z
Imada, Y., Tatebe, H., Ishii, M., Chikamoto, Y., Mori, M., Arai, M., et al. (2015). Predictability of two types of El Niño assessed using an extended seasonal prediction system by MIROC. Monthly Weather Review, 143, 4597–4617. https://doi.org/10.1175/mwr-d-15-0007.1

Ineson, S., & Scaife, A. A. (2009). The role of the stratosphere in the European climate response to El Niño. Nature Geoscience, 2(1), 32–36. https://doi.org/10.1038/ngeo381

Inoue, J., Hori, M. E., & Takaya, K. (2012). The role of Barents sea ice in the wintertime cyclone track and emergence of a warm-Arctic cold-Siberian anomaly. Journal of Climate, 25(7), 2561–2568. https://doi.org/10.1175/JCLI-D-11-00449.1

Ishii, M., & Kimoto, M. (2009). Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. Journal of Oceanography, 65(3), 287–299. https://doi.org/10.1007/s10872-009-0207-7

Ishii, M., Kimoto, M., Sakamoto, K., & Iwasaki, S. I. (2006). Stratic sea level changes estimated from historical subsurface temperature and salinity analyses. Journal of Oceanography, 62(2), 155–170. https://doi.org/10.1007/s10872-006-0041-y

Jin, E., Kinter, J., Wang, B., Park, C. K., Kang, I. S., Kirtman, B., et al. (2008). Current status of ENSO prediction skill in coupled atmosphere models. Geophysical Research Letters, 35, 1–6. https://doi.org/10.1002/gRL.20085

Jin, E., Kinter, J., Wang, B., Park, C. K., Kang, I. S., Kirtman, B., et al. (2008). Current status of ENSO prediction skill in coupled atmosphere models. Geophysical Research Letters, 35, 1–6. https://doi.org/10.1002/gRL.20085

Jin, E., Kinter, J., Wang, B., Park, C. K., Kang, I. S., Kirtman, B., et al. (2008). Current status of ENSO prediction skill in coupled atmosphere models. Geophysical Research Letters, 35, 1–6. https://doi.org/10.1002/gRL.20085

Kabasawa, I., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan, 93, 5–48. https://doi.org/10.2151/jmsj.2015-001

Kamuro, Y. (2014). The impact of surface mixing on the Arctic river water distribution and stratification in a global ice-ocean model. Journal of Climate, 27, 4359–4370. https://doi.org/10.1175/JCLI-D-13-00090.1

Katsu, Y., & Xie, S.-P. (2013). Recent global-warming hiatus tied to equatorial Pacific surface cooling. Nature, 501(7467), 403–407. https://doi.org/10.1038/nature12534

Katsu, Y., & Xie, S.-P. (2016). The tropical Pacific as a key pacemaker of the variable rates of global warming. Nature Geoscience, 9, 669–673. https://doi.org/10.1038/ngeo2770

Kushnir, Y., Scalef, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., et al. (2019). Towards operational predictions of the near-term climate. Nature Climate Change, 9, 94–101. doi:10.1038/s41558-018-0359-7

Luo, J.-J., Masson, S., Behera, S., & Yamagata, T. (2008). Extended ENSO predictions using a fully coupled ocean–atmosphere model. Journal of Climate, 21, 84–93. https://doi.org/10.1175/JCLI-D-11-00121.2

Madden, R. A., & Julian, P. R. (1972). Description of global-scale circulation cells in the tropics with a 40-50 day period. Journal of the Atmospheric Sciences, 29(6), 1109–1122. https://doi.org/10.1175/1520-0469(1972)029<1109:DSGSCC>2.0.CO;2

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society, 78, 1069–1079. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOC>2.0.CO;2

Marshall, A. G., Hendon, H. H., Feng, M., & Schiller, A. (2015). Initiation and amplification of the Ningaloo Niño. Climate Dynamics, 45(9-10), 2367–2385. https://doi.org/10.1007/s00382-015-2477-5

Matsumo, T. (1966). Geostrophic motions in the equatorial area. Journal of the Meteorological Society of Japan, 44(1), 25–43. https://doi.org/10.2151/jmsj1965.44.1_25

Meeth, G. A., Hu, A., & Huiyan, T. (2016). Initialized decadal prediction for transition to positive phase of the Interdecadal Pacific Oscillation. Nature Communications, 7, 17718. https://doi.org/10.1038/ncomms1718

Meeth, G. A., & Teng, H. (2012). Case studies for initialized decadal hindcasts and predictions for the Pacific region. Geophysical Research Letters, 39, L22705. https://doi.org/10.1029/2012GL053423

Meeth, G. A., & Teng, H. (2014). CMIP5 multi-model initialized decadal hindcasts for the mid-1970s shift and early-2000s hiatus and predictions for 2016–2035. Geophysical Research Letters, 41, 1711–1716. https://doi.org/10.1002/2014GL059528

Merryfield, W. J., Baehr, J., Battè, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., et al. (2020). Current and emerging developments in subseasonal to decadal prediction. Bulletin of the American Meteorological Society, 101, E869–E896. https://doi.org/10.1175/BAMS-D-19-0037.1

Mochizuki, T., Chikamoto, Y., Kimoto, M., Ishii, M., Tatebe, H., Komuro, Y., et al. (2012). Decadal prediction using a recent series of MIROC global climate models. Journal of the Meteorological Society of Japan, 90, 373–383. https://doi.org/10.2151/jmsj.2012-A22

Mochizuki, T., Ishii, M., Kimoto, M., Chikamoto, Y., Watanabe, M., Nozawa, T., et al. (2010). Pacific decadal oscillation hindcasts relevant to near-term climate prediction. Proceedings of the National Academy of Sciences of the United States of America, 107(5), 1833–1837. https://doi.org/10.1073/pnas.0906531107

Mochizuki, T., Tatebe, H., Kataoka, T., & Koyama, H. (2019a). MIROC MIROC6 model output prepared for CMIP6 DCPP dcppA-assim. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5467

Mochizuki, T., Tatebe, H., Kataoka, T., & Koyama, H. (2019b). MIROC MIROC6 model output prepared for CMIP6 DCPP dcppA-hindcast. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5468

Mori, M., Kosaka, Y., Watanabe, M., Nakamura, H., & Kimoto, M. (2019). A reconciled estimate of the influence of Arctic sea-ice loss on recent Eurasian cooling. Nature Climate Change, 9, 123–128. https://doi.org/10.1038/s41558-018-0379-3

KATAOKA ET AL.
Sudo, K., Takahashi, M., Kurokawa, J., & Akimoto, H. (2002). CHASER: A global chemical model of the troposphere 1. Model description. Journal of Geophysical Research, 107(D17), 4339. https://doi.org/10.1029/2001jd001113

Tatebe, H., Ishii, M., Mochizuki, T., Chikamoto, Y., Sakamoto, T. T., Komuro, Y., et al. (2012). The initialization of the MIROC climate models with hydrographic data assimilation for decadal prediction. Journal of the Meteorological Society of Japan, 90A, 275–294. https://doi.org/10.2151/jmsj.2012-A14

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Oguchi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geoscientific Model Development, 12, 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Tatebe, H., & Watanabe, M. (2018). MIROC MIROC6 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5603

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experimental design. Bulletin of the American Meteorological Society, 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tomkins, A. M., Ortiz De Zarate, M. I., Saurral, R. I., Vera, C., Saulo, C., Merry, A. M., & Ortiz, J. (2019). PDES: A high resolution decadal hindcast prepared for CMIP6. Geoscientific Model Development, 12, 3149–3197. https://doi.org/10.5194/gmd-12-3149-2019

Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., et al. (2018). Predicting near term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. Bulletin of the American Meteorological Society, 99, 1867–1886. https://doi.org/10.1175/BAMS-D-17-0098.1

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late 20th century North Atlantic Ocean heat content. Journal of Climate, 25, 5173–5189. https://doi.org/10.1175/JCLI-D-11-0595.1

Zhang, L., & Han, W. (2018). Impact of Ningaloo Niño on tropical pacific and an interbasin coupling mechanism. Geophysical Research Letters, 45, 1392–1398. https://doi.org/10.1002/2016GL067762

Zhang, L., Han, W., Li, Y., & Shinoda, T. (2018). Mechanisms for generation and development of Ningaloo Niño. Journal of Climate, 31, 9239–9259. https://doi.org/10.1175/JCLI-D-18-0175.1