A DRP-4DVar-Based Coupled Data Assimilation System With a Simplified Off-Line Localization Technique for Decadal Predictions

Yujun He1,2, Bin Wang1,2,3,4, Li Liu2,4, Wenyu Huang2,4, Shiming Xu2,4, Juanjuan Liu1,3, Yong Wang2,4, Lijuan Li1, Xiaomeng Huang2,4, Yiran Peng2,4, Yanluan Lin2,4, and Yongqiang Yu1,3

1State Key Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, 2Ministry of Education, Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua University, Beijing, China, 3College of Ocean Science, University of Chinese Academy of Sciences, Beijing, China, 4Joint Centre for Global Change Studies, Beijing, China

Abstract A new weakly coupled data assimilation (CDA) system based on the dimension-reduced projection four-dimensional variational data assimilation (DRP-4DVar) with a simplified off-line localization technique and a fully coupled model, i.e., the Grid-point Version 2 of Flexible Global Ocean-Atmosphere-Land System Model (FGOALS-g2), was developed for the initialization of decadal predictions. A 1-month assimilation window was adopted for the CDA system, in which monthly mean temperature and salinity analyses were assimilated along the trajectory of the coupled model during the initialization for the period of 1945–2006. The system is efficient because the 62-year initialization only takes about 2.375 times of the time cost of the uninitialized run for the same period. Compared with the uninitialized simulation and the initialization without localization, ocean temperature and salinity, sea surface elevation, surface air temperature, and precipitation are in better agreement with the verification data. Furthermore, climate variabilities in the Pacific and Atlantic regions such as El Niño-Southern Oscillation, Pacific Decadal Oscillation (PDO), and Atlantic Multidecadal Oscillation (AMO) are more realistically captured. Starting from the initial conditions (ICs) generated by the initialization, 10-member ensemble decadal prediction experiments were conducted each year from 1961 to 1996. The results demonstrate that higher decadal prediction skills of surface air temperature anomalies averaged over the globe, ocean, land, and the North Pacific subpolar gyre are achieved than those obtained from persistence, the uninitialized simulation, and the prediction initialized from the ICs without localization. Besides, PDO and AMO indices exhibit significant correlation skills in most lead times.

Plain Language Summary The decadal prediction is important for medium- and long-term planning in various fields, e.g., infrastructure, agriculture, fishery, electric power, and tourism. It is usually made by a climate model which includes the ocean, atmosphere, land, and sea ice components. The decadal prediction is initialized from a close-to-observed initial state which consistently combines observations and model simulations using a kind of methods called “data assimilation (DA).” However, DA methods for initializing decadal predictions currently around the world are still at an exploring stage, which leads to the inconsistency between the model and the initial state, possibly degrading the prediction skill. This study focuses on proposing a novel DA method for initializing decadal predictions, which is time-saving and produces initial states consistent with the model. This method is able to well reproduce the multiyear mean ocean temperature and salinity, sea surface elevation, surface air temperature, and precipitation accurately. It also accurately represents the notable climate oscillations in the Pacific and Atlantic. The decadal prediction initialized by the new method shows high skills in presenting surface air temperature and the climate oscillations in the Pacific and Atlantic.

1. Introduction

Decadal predictions of the next 10–30 years can benefit the medium- and long-term decision-making strategies of various industries (e.g., agriculture, fisheries, water conservancy, hydropower, shipping, and tourism) in a country (Crawford et al., 2006). Decadal predictions using coupled models or earth system models have been a rapidly developing field in the recent decade. It is included as one of the core experiments in the fifth
The coupled model used in the initialization and decadal prediction experiments is FGOALS-g2. Model Description

A majority of modeling centers around the world choose coupled data assimilation (CDA) methods during initialization, which applies the data assimilation technique to one or several components of the coupled model, and a long-term assimilation cycle is conducted under the framework of the coupled model. Most of these methods use a simple nudging (Keenlyside et al., 2008; Merryfield et al., 2013; Sanchez-Gomez et al., 2016; Smith et al., 2013; Swingedouw et al., 2013; Wei et al., 2017; Xin et al., 2013) or a nudging-based Incremental Analysis Update (IAU) approach (Merryfield et al., 2013; Tatebe et al., 2012; Wu et al., 2015, 2018). A small number of models apply more advanced techniques, including variational methods such as three- or four-dimensional variational (3DVar or 4DVar) data assimilation (Mochizuki et al., 2016; Saha et al., 2010; Wang et al., 2013) and ensemble techniques, such as the ensemble optimal interpolation (EnOI) (Ham et al., 2014; Wu et al., 2018) or ensemble Kalman filter (EnKF) (Zhang et al., 2007). Among these methods, nudging, IAU, 3DVar, and EnOI are time-saving. However, the ICs obtained using the first three methods are less consistent with the coupled model than the ensemble methods. The ICs obtained from 4DVar tend to have a better consistency with the model than simpler assimilation methods like nudging, optimal interpolation, and 3DVar, as 4DVar provides a good fit to observations along the model trajectory in the assimilation window due to the incorporation of model constraint (Kalnay, 2003; Sugiura et al., 2008). However, the computational cost of 4DVar is extremely expensive, owing to the computation of the adjoint model. Apart from the CDA methods, a minority of models adopt uncoupled initialization methods that use the results of the stand-alone model forced with observations or reanalysis data as ICs (Bellucci et al., 2013; Du et al., 2012; Karspeck et al., 2015; Troccoli & Palmer, 2007; Yeager et al., 2012). Unfortunately, the ICs obtained from the uncoupled methods result in worse consistency between their model components than the CDA approaches (Balmaseda et al., 2009).

Recently, He et al. (2017) applied a 4DEnVar assimilation method, i.e., the dimension-reduced projection four-dimensional variational data assimilation (DRP-4DVar) (Wang et al., 2010) to a coupled model, i.e., the Flexible Global Ocean–Atmosphere–Land System Model, Grid-point Version 2 (FGOALS-g2) (Li et al., 2013). Unlike the previously described initialization methods, which are either consistent with model dynamics and physics or time-saving, the DRP-4DVar method can achieve both goals. In their work, the initial shock during decadal predictions is remarkably reduced and is one of the smallest among the CMIP5 models. These authors also achieved a high decadal prediction skill of the global mean surface air temperature anomaly (SATA). However, the performance of the DRP-4DVar-based CDA system has not been systematically explored. In this study, the performance of the CDA system in terms of the observational cost function, bias, root mean square error (RMSE), climate mean states, and climate variabilities is examined. Also, a more detailed evaluation of the decadal prediction skills is conducted.

The paper is organized as follows. An introduction of the DRP-4DVar-based CDA system and experiments is presented in section 2. A comprehensive assessment of the CDA system is given in section 3. In section 4, the decadal prediction skills are discussed before major conclusions are drawn in section 5.

2. Model, CDA System, and Experiments

2.1. Model Description

The coupled model used in the initialization and decadal prediction experiments is FGOALS-g2, which was jointly developed by the Chinese Academy of Sciences and Tsinghua University and participated in CMIP5 (Li, et al., 2013). The atmospheric component is GAMIL2 (Li et al., 2013), which employs a Gaussian grid between 65.58°S and 65.58°N with a 2.8° spacing and a weighted equal-area grid poleward of 65.58° with
an interval larger than 2.8° in the horizontal direction and has 26 vertical σ levels extending to 2.194 hPa. The ocean component is LICOM2 (Liu et al., 2012), whose horizontal resolution is 1°×1° but increases to 0.5° in the tropics in the meridional direction. This component includes 30 vertical layers with an interval of 10 m in the upper 150 m. The sea ice component is CICE4-LASG, which has an approximately identical horizontal resolution to that of the ocean component (Yu et al., 2008). The land component is CLM3, which has a horizontal resolution that is the same as that of the atmospheric component (Oleson et al., 2010). These four components exchange fluxes through the coupler CPL6 (Craig et al., 2005). The coupling frequency is 1 day for the ocean component and 1 hr for the other components.

2.2. DRP-4DVar-Based CDA System

The DRP-4DVar-based CDA system assimilates the full-field monthly mean ocean temperature and salinity analyses (i.e., ds285.3) of the upper 1,000 m (Ishii et al., 2005, 2006) into the ocean component of FGOALS-g2 (He et al., 2017). The system has the following advantages. On account of the coupled model constraint like 4DVar, it can generate consistent ICs with the coupled model, which best fits ds285.3 analyses along the trajectory of the coupled model in 1-month assimilation windows. Moreover, it is cost-effective due to the abandonment of the adjoint model and a weak flow-dependent 30-member ensemble, which remarkably reduces the time expense compared to 4DVar and EnKF, respectively. The 30 members are selected from a fixed data set of the 100-year Pre-Industrial Control (PI-CTRL) run in terms of their correlations with corresponding observational innovations, whose coefficients are the largest so that these members change from month to month. Thus, the CDA system eliminates the need to integrate the coupled model 30 times simultaneously as the strong flow-dependent ensemble in 4DEnVar. As the ensemble size is far smaller than the dimension of both model and ds285.3 analyses, localization is a necessity to reduce the sampling error and the spurious long-distance correlations in the background error covariance as well as increase the rank of the ensemble by limiting the impact of ds285.3 on the analysis at a particular grid point (Hamill et al., 2001; Lorenc, 2003; Oke et al., 2007, 2010; Wang et al., 2018). For the sake of low computing expense, a simplified off-line localization scheme, which was based on the method proposed by Wang et al. (2018) and applied to the coupled model for the first time, is adopted in the DRP-4DVar-based CDA system. The initialization with this CDA system consists of two steps and is illustrated in Figure 1.

The first step is the DRP-4DVar-based assimilation without localization, which aims to obtain ocean analysis of the first day in each month (i.e., IC) and monthly mean analysis without localization. This step requires three inputs, including 30 perturbation samples generated from the PI-CTRL run, the background (x_{0,1}), and observational innovation (y_{obs}), i.e., the difference between the ds285.3 analyses (y_{obs}) and the observation-space background (y_{b}). The 30 perturbation samples consist of 30 monthly mean perturbation samples and 30 IC perturbation samples and are obtained as follows. First, we select 100 years from the FGOALS-g2 PI-CTRL run when it reaches an equilibrium state, acquiring 1,200 monthly mean samples and 1,200 IC samples (i.e., the first day of each month). Then, we gain 1,200 monthly mean perturbation samples and 1,200 IC perturbation samples by subtracting the corresponding average of the 100 years in each month, aiming to minimize the impact of external forcings. Each month, 30 monthly mean perturbation samples from the total 100 samples are selected using a successive regression method to ensure the sample independence. The first monthly mean perturbation sample is selected with the largest absolute value of correlation with the observational innovation. The second monthly mean perturbation sample is selected with the largest absolute value of correlation with the residual, which is computed as the difference between the observational innovation and the projection of the first selected monthly mean perturbation sample onto the observational innovation. The other 28 monthly mean perturbation samples are selected in the same way as the second one except that the residual is computed as the difference between the residual obtained in the last selected sample and the projection of the last selected monthly mean perturbation sample onto that residual. Lastly, the corresponding 30 IC perturbation samples are obtained.

For each month, a 1-month free coupled model integration is conducted firstly to obtain the background (y_{b}). Then, an analysis (x_{a,1}) of ocean temperature, salinity, sea surface height (SSH), and seawater velocity can be acquired at the beginning of the month using the DRP-4DVar method. Afterward, a second 1-month free model integration is started using the ocean analysis and the same ICs of the other four components as the first 1-month free model integration to obtain the background of the next month as well as the monthly
mean analysis after assimilation without localization ($x_{a,1}$). This process is cycled for 62 years from 1945 to 2006.

The second step is a simplified off-line localization, which aims to acquire final ocean analysis of the first day in each month (i.e., IC) after localization ($x_a$). Here, we adopt an economic localization scheme, which expands the correlation function of the localization matrix based on a limited number of principal modes (Wang et al., 2018) without running the coupled model (i.e., off-line). Similar to the method of assimilating observations one by one in the serial processing method of EnKF, the principal mode is used one by one owing to the orthogonality between the principal modes. As Gaspari and Cohn (1999) show that their correlation function is very similar to the form of the exponential function, the correlation function can be separable in the zonal, meridional and vertical directions, making it possible to implement localization in each direction separately. As Wang et al. (2018) pointed out, a minimum of 20 principal modes constitute the dominant part of the correlation function, accounting for more than 97% of the total modes in the one-dimensional case. Further taking into account the different lengths of the three directions and less time expense, 60, 30, and 20 principal modes are selected in the zonal, meridional and vertical directions, respectively.

For each month, the localization is performed iteratively in the three directions. In the zonal direction, the background is updated as the analysis in the first step ($x_{b,2} = x_{a,1}$). Sixty extended samples are produced by the Schur products between the 60 principal modes and one analysis increment sample $x_{b,1}$, which is the analysis increment obtained in the first step. Then, a second analysis increment $x_{c,2}$ is acquired using DRP-4DVar. A second analysis is completed ($x_{a,2} = x_{b,2} + x_{c,2}$). In the meridional direction, the background is updated as the second analysis ($x_{b,3} = x_{a,2}$). Thirty extended samples are generated by the Schur products between 30 principal modes and one analysis increment sample $x_{b,2}$. A third analysis increment $x_{c,3}$ is acquired using DRP-4DVar. A third analysis is obtained ($x_{a,3} = x_{b,3} + x_{c,3}$). The same process is applied along the vertical direction with the updated background ($x_{b,4} = x_{a,3}$), 20 extended samples and one analysis increment sample $x_{c,3}$ to obtain the fourth analysis increment $x_{c,4}$ using DRP-4DVar. The fourth analysis is then acquired ($x_{a,4} = x_{b,4} + x_{c,4}$), which is the final analysis after localization ($x_a = x_{a,4}$).

The time expense of the CDA system is approximately 19 min altogether for each month, which is 2.375 times of the time cost of the uninitialized (UNINIT) run (i.e., 8 min). The first step takes about 17.3 min, which includes two 1-month free model integrations with 108 central processing units (CPUs) and 1.3-min-long DRP-4DVar assimilation with 1 CPU. The second step costs 1.5 min with 1 CPU.
To compare the effect of localization on the initialization results and obtain the monthly mean analysis after localization (\(y_a\)), another 1-month free model integration is conducted, initialized from the final ocean analysis produced and the ICs of the other four components of the coupled model in the first step.

The 1-month free model integration starting from the analysis or final analysis can check whether the monthly mean assimilated ocean states best fit ds285.3 analyses along the trajectory of the coupled model. Meanwhile, the other components are adjusted under the constraints of the coupled model during the 1-month integration.

### 2.3. Experiment Design

Starting from the ICs from the second member of the UNINIT experiment (i.e., the historical run) from the FGOALS-g2 in CMIP5, one initialization experiment without localization (INIT_NOLOC), one initialization experiment with localization (INIT), and one UNINIT experiment are conducted from 1945 to 2006. Using the ICs from the INIT experiment, 37 sets of 10-member ensemble 10-year hindcast experiments (HCST) are carried out every year from 1961 to 1997. Another set of hindcast experiments (HCST_NOLOC) starting from the INIT_NOLOC run are conducted the same way as the HCST run. The external forcings of all the experiments are identical, which are the observed forcings before 2005 and the Representative Concentration Pathway 4.5 (RCP4.5) from 2006 onward (Taylor et al., 2012). The ICs of the 10 members are generated using a time-lagged method, i.e., 1 February to November each month of the year before each starting year.

Data used for evaluating the validity and skill of the CDA system are EN.4.2.1 analyses for ocean temperature and salinity (Good et al., 2013); Hadley Centre Sea Ice and Sea Surface Temperature (HadISST1.1) analysis for sea surface temperature (SST) (Rayner et al., 2003); ORAS4 reanalysis (Balmaseda et al., 2013) and SODA 2.2.4 reanalysis (Carton & Giese, 2008) for sea surface salinity (SSS) and SSH; HadCRUT4 analysis (Morice et al., 2012) for SATA; Global Precipitation Climatology Project (GPCP) version 2.3 analysis (Adler et al., 2003) for precipitation; and NCEP/NCAR Reanalysis 1 for sea level pressure (SLP), zonal wind, and 500-hPa geopotential height (Kalnay et al., 1996).

### 3. Assessment of the CDA System

#### 3.1. Bias and RMSE of Ocean Temperature and Salinity

The area-weighted bias and RMSE, which represent systematic and random errors as those defined in Taylor (2001), Gleckler et al. (2008) and Goddard et al. (2013), are essential metrics used to assess the accuracy of the CDA system and are respectively defined as

\[
\text{bias} = \frac{\sum_{i=1}^{n} \text{diff}(i) \times \text{area}(i)}{\sum_{i=1}^{n} \text{area}(i)},
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\text{diff}(i) - \text{bias})^2 \times \text{area}(i)}{\sum_{i=1}^{n} \text{area}(i)}},
\]

where \(\text{diff}(i)\) denotes the difference between simulated and observed ocean temperature or salinity of grid point \(i\), and \(\text{area}(i)\) stands for the area of grid point \(i\).

The ocean temperature of the UNINIT run shows a cold bias in the upper 150 m and below 700 m and a warm bias in the 150–700-m range (Figure 2a). The temperature analyses in both INIT_NOLOC and INIT runs reduces the bias in the upper 600 m. Compared to the temperature RMSE of the UNINIT run, the RMSE of the temperature analysis in the INIT_NOLOC run is much smaller above 400 m and slightly larger below 400 m (Figures 2b). The INIT run shows a more reduction of the temperature RMSE in the upper 400 m than the INIT_NOLOC run, probably due to the localization adopted in the INIT run. The slightly increased temperature RMSE below 400 m in both the INIT_NOLOC and INIT runs is probably due to the overestimation of ocean observational error. The salinity of the UNINIT run shows a negative bias in the upper 800 m and a positive bias below 800 m (Figures 2c). The salinity analyses in the INIT_NOLOC and INIT runs reduce the negative bias above 600 m (Figure 2c). Compared with the UNINIT run, the RMSE of the salinity analyses in the INIT_NOLOC and INIT runs is larger in the 100–500-m range (Figure 2d). The small difference between the INIT_NOLOC and INIT runs indicates that the effect of the
localization in the INIT run is very limited, which calls for further improvement. Some degradations are also found in the deep ocean. They may be probably caused by the inaccurate estimation of observational error, imbalances between the ocean variables during the assimilation and the quality of the assimilated data set, which requires further improvement of the CDA system.

### 3.2. SST, SSS, and SSH

The temporal evolution of the annual global mean SST anomaly (SSTA) analysis in the INIT run corresponds well with HadISST, showing a correlation of 0.91, higher than that in the UNINIT simulation (0.84) (Figure 3a). The linearly detrended SSTA analysis in the INIT run shows more accurate internal variability, with a correlation of 0.77, than that in the UNINIT run (0.40) (Figure 3b). The INIT_NOLOC run shows comparable performance with the INIT run. These results indicate that the accuracy of the global mean SSTA analysis relies not only on external forcings but also on initialization.

Global distribution of the climatological mean of assimilated SST (Figures 4c and 4d) agrees better with HadISST (Figure 4a) in the tropics than that of the UNINIT run (Figure 4b). The RMSEs averaged over the tropics are 0.97°C and 0.94°C for the INIT_NOLOC and INIT runs, respectively, both smaller than those for the UNINIT run (1.13°C). The assimilated SST presents a larger warm pool in the western Pacific and the Indian Ocean and a more westward extension of the cold tongue in the eastern Pacific than those in the UNINIT run. In particular, the analyses of both INIT_NOLOC and INIT runs reduce the cold biases along and to the east of several western boundary currents (i.e., Kuroshio, East Australian Current, Gulf Stream, and Brazil Current) and the warm biases along the west coasts of South America and Africa as well.
as in the Labrador Sea (Figures 4e–4i). The INIT run shows overall smaller biases in the above-mentioned areas than the INIT_NOLOC run (Figure 4j).

The correlation of the annual mean SST with HadISST is shown in Figure 5. The uninitialized SST shows significant positive correlations in some parts of the tropical Indian Ocean, tropical western and central Pacific, and tropical and midlatitude Atlantic (Figure 5a), while few areas have significant correlations after linearly detrending (Figure 5e). The SST analyses of the INIT_NOLOC and INIT runs (Figures 5b and 5c) present higher correlations in tropical oceans and some parts of the midlatitude ocean than those of the UNINIT run. After linearly detrending, the analyzed SST variability (Figures 5f and 5g) still has high correlations in the tropical Pacific, midlatitude North Pacific, and the northern part of the North Atlantic, indicating that the improved SST accuracy of these regions is dominated by initialization rather than external forcings. Additionally, the SSTs in the tropical Indian Ocean, tropical western Pacific, and tropical Atlantic also have significant correlations with observed but spanning fewer areas than those before detrending, demonstrating that the SST accuracy over these regions depends on the combined response to initialization and external forcings. Compared with the INIT_NOLOC run (Figures 5d and 5h), most areas around the globe in the INIT run present higher correlations, indicating the localization is likely to improve the accuracy of the assimilation. However, some scattered areas show less

Figure 3. (a) Time series of annual global mean SSTA and (b) linearly detrended SSTA from 1955 to 2006 (unit: °C). The black, green, blue, and red lines denote HadISST, UNINIT, INIT_NOLOC, and INIT runs, respectively. The SSTA correlations with HadISST are given in parentheses. The anomalies are relative to the climatology from 1961 to 1990.

Figure 4. Spatial distribution of (a–d) SST and (e–j) SST difference averaged from 1955 to 2006. (a)–(d) represent SST spatial distribution from (a) HadISST, (b) UNINIT experiment, (c) INIT_NOLOC run, and (d) INIT run. (e)–(j) stand for SST spatial difference between (e) the UNINIT experiment and HadISST, (f) the INIT_NOLOC run and HadISST, (g) the INIT run and HadISST, (h) the INIT_NOLOC and UNINIT runs, (i) the INIT and UNINIT runs, and (j) the INIT and INIT_NOLOC runs.
accuracy, which calls for further study and improvement of the localization scheme.

The correlation of the detrended SST analysis in the tropics is higher than that in the extratropical regions, which agrees with Wu et al. (2018). A possible explanation for this is that the atmosphere is primarily forced by the ocean in the tropics (Gill, 1980; Lindzen & Nigam, 1987); conversely, the ocean is mainly forced by the atmosphere in extratropical regions (Frankignoul & Hasselmann, 1977). Here, only ocean observations are assimilated to change only ocean state variables at the start of each month; other model components are adjusted or constrained in the coupled simulation in the 1‐month time window. The tropical SST responds well to the observationally constrained ocean state, while the SSTs in the extratropical areas also rely on the air‐sea interactions of the coupled model. It is speculated that incorporating the ocean observations into atmospheric variables through considering the air‐sea covariance of background error during the assimilation analysis process or consistently assimilating atmospheric observations in the meantime may further enhance the SST correlation in the extratropical regions. The SST correlation in the tropical Pacific is the highest, which is also consistent with Wu et al. (2018). The reason for this pattern is that tropical Pacific SST is primarily controlled by self‐sustaining internal dynamics (Chen et al., 2004), while the SSTs in the North Pacific, tropical Atlantic and the Indian Ocean have a close link with the SST in the tropical Pacific through the atmospheric teleconnection (Alexander et al., 2002). Therefore, the tropical Pacific SST largely depends on ocean assimilation, while the SSTs in other regions also rely on the model's ability to simulate the atmospheric teleconnection.

The time series of the annual global mean SSS anomaly (SSSA) are shown in Figure 6. The SSSA analyses of both the INIT_NOLOC and INIT runs are closer to ORAS4 and SODA reanalyses than that of the UNINIT run, with higher correlations not only with trend but also without trend. However, large disparities exist in the two SSS reanalyses data sets. Possible reasons include the sparseness of salinity observations, different models and assimilation methods (Balmaseda et al., 2013; Carton & Giese, 2008). The inaccurate SSSA analyses of the INIT_NOLOC and INIT runs may possibly due to the incorrect assimilated ds285.3 data set. Global distributions of the climatological mean of assimilated SSS in the INIT_NOLOC and INIT runs reduce the negative biases in most parts of the Pacific, South Atlantic, and northern Indian Ocean and the positive bias in the Labrador Sea (Figure 7). The localization only has a small
contribution to the accuracy of the SSS spatial distribution (Figure 7j). The highest SSS correlation from the initialization is located in the tropical Pacific (Figure S1 in supporting information), which agrees with the SST results (Figure 5).

Global distributions of the climatological mean SSH are shown in Figure 8. The assimilated SSH in the INIT_NOLOC and INIT runs presents a more accurate zonal distribution in the Pacific and Indian Ocean than that in the UNINIT run (Figures 8a–8d). Besides, it corrects the higher-than-observed SSH in the tropical Indian Ocean and Pacific and the lower-than-observed SSH along the east coast of America (Figures 8e–8i). However, the localization fails to show much improvement of the SSH spatial distribution (Figure 8j).

3.3. SATA and Precipitation

The assimilated global mean SATA (Figures 9a and 9b) is in better agreement with HadCRUT4 than that of the UNINIT run. The linearly detrended SATA analyses of the INIT_NOLOC and INIT runs are more advantageous, with correlations of 0.72 and 0.76, respectively, than that of the UNINIT run (0.51). Thus, the high accuracy of the SATA analysis results from both external forcings and initialization. The SATA analyses averaged above the ocean (Figures 9c and 9d) show similar results with the SSTA analyses (Figure 3). The land mean SATA analysis of the INIT_NOLOC run has similar performance with the UNINIT run (Figure 9e), which mostly results from the increasing trend. After linearly detrending, both the INIT_NOLOC and INIT runs show much better performances than the UNINIT run (Figure 9f). Besides, the INIT run is more advantageous than the INIT_NOC run not only with trend but also without trend, which is possibly owing to localization that reduces the long-distance spurious correlation to obtain a more accurate ocean analysis. The more accurate assimilated ocean state can then influence the atmospheric...
variables over land through the air-sea interactions and dynamic processes in the atmosphere (e.g., advection and teleconnection) of the coupled model in the 1-month time window, thereby producing a more accurate atmospheric state over land. In addition, the correlation of assimilated ocean mean SATA is higher than that of the land, indicating that the assimilated ocean mean SATA is directly influenced by the underlying ocean through the model air-sea interactions, while the land mean SATA analysis is further affected by the dynamic processes in the atmosphere, which allows the propagation of the observational information in the ocean state variables to the SATA over land in the 1-month time window.

The tropical precipitation of GPCP (Figure 10a) exhibits several maxima in the northern tropical Pacific, which is parallel to the equator; the southern tropical Pacific, which is oriented in the northwest-southeast direction (known as the South Pacific Convergence Zone, SPCZ); the tropical Indian Ocean; and a dry zone to the west coast of South America. This distribution corresponds to the locations of the warm pools and cold tongue SST patterns (Figure 4a). The UNINIT run (Figure 10b) shows more precipitation in the tropical Pacific and less rainfall in the tropical eastern Indian Ocean. The SPCZ lies almost parallel to the equator and extends more eastward than GPCP (i.e., double ITCZ), which is similar to the corresponding SST distribution (Figure 4b). The dry zone extends less westward than GPCP, which agrees with the reduced westward extension of the cold tongue (Figure 4b). The assimilated precipitation in the INIT run (Figure 10d), with similar performance with that in INIT_NOLOC run (Figures 10c and 10j), shows smaller bias and RMSE (0.30 and 1.44 mm/day) than that in the UNINIT run (0.35 and 1.57 mm/day) in the tropics. In particular, both INIT_NOLOC and INIT runs alleviate the eastward extension of the SPCZ and the reduced westward extension of the dry zone in the UNINIT run and present more rainfall in the tropical eastern Indian Ocean than the UNINIT run (Figures 10c–10i), which are in better agreement with GPCP.

Figure 8. As in Figure 4, but for SSH. The referenced data set is ORAS4 reanalysis.
3.4. El Niño–Southern Oscillation

One of the most prominent interannual variabilities is El Niño-Southern Oscillation (ENSO), which has a profound influence on the global climate through atmospheric teleconnections (Alexander et al., 2002; Pan & Oort, 1983). ENSO is a product of air-sea interactions in the tropics, usually characterized by anomalously warm (El Niño) or cold (La Niña) SSTs in the central and eastern equatorial Pacific occurring every 2–7 years, associated with SLP seesaws between the Indian Ocean and the Pacific (Southern Oscillation) (McPhaden et al., 2006). El Niño and La Niña are usually monitored by SSTA averaged in the Niño 3.4 region (5°S–5°N, 120°W–170°W), i.e., the Niño 3.4 index. The time evolutions of the assimilated monthly Niño 3.4 index in the INIT_NOLOC and INIT runs exhibit higher correspondence with HadISST (correlation is 0.92) than that of the UNINIT run (0.19; see Figure 11a). However, the amplitudes of several El Niño and La Niña events are smaller than those observed, which can also be demonstrated by the smaller standard deviation of Niño 3.4 index (Figure S2 in supporting information).

The assimilated monthly Southern Oscillation Index (SOI) and zonal wind stress anomaly (ZWSA) averaged in the Niño 3.4 region in the INIT run (Figures 11b and 11d) correspond better with NCEP reanalysis (the correlation is 0.37 for SOI and 0.44 for ZWSA) than those of the UNINIT run (0.09 and 0.13, respectively, which are not shown in Figures 11b and 11d). The INIT_NOLOC run (not shown in Figures 11b and 11d) performs better for SOI (correlation is 0.41) and worse for ZWSA (correlation is 0.41) than the INIT run. Furthermore, the assimilated monthly SOI and ZWSA are not as good as the assimilated monthly Niño 3.4 index (Figure 11a). One cause is that only the ocean variables in the ICs are changed at each assimilation step. Unlike the Niño 3.4 index, which directly depends on initialization, SOI and ZWSA hinge on the air-sea interactions of the coupled model during the 1-month integration in the assimilation window. The other cause may be the noise in the atmosphere at short time scales. After performing the 37-month running mean of SOI and ZWSA to reduce the noise (Figures 11c and 11e), the INIT_NOLOC and INIT runs agree more with NCEP with larger correlations than those at the monthly time scale.

The Niño 3.4 index is closely linked with SOI and ZWSA with respect to the verification data, with correlations of −0.72 and 0.64, respectively, implying that during El Niño (La Niña), the positive (negative) Niño 3.4

Figure 9. Time series of annual global, ocean, and land mean SATA (left) and linearly detrended SATA (right) from 1955 to 2006. The black, green, blue, and red lines denote HadCRUT4, UNINIT experiment, INIT_NOLOC, and INIT runs, respectively. The SATA correlations with HadCRUT4 are given in parentheses. The anomalies are relative to the climatology from 1961 to 1990.
The index is accompanied by negative (positive) SOI and eastward (westward) ZWSA. The lower-than-observed amplitude of assimilated Niño 3.4 index is associated with larger-than-observed assimilated SOI and lower-than-observed assimilated ZWSA of several major El Niño events in 1957–1958, 1965–1966, 1972–1973, 1982–1983, and 1986–1988, as well as smaller-than-observed assimilated SOI and higher-than-observed assimilated ZWSA of major La Niña events in 1955–1956, 1964–1965, 1970–1971, and 1988–1989.

### 3.5. PDO, AMO, and AMOC

The Pacific Decadal Oscillation (PDO) is a dominant decadal climate variability mode in the North Pacific comprising SST anomalies in the midlatitude North Pacific and those in its surrounding areas showing opposite changes. The PDO index is defined as the normalized time series of the leading empirical orthogonal function mode for the linearly detrended SSTA in the North Pacific (20°–70°N, 110°–260°E) (Mantua et al., 1997). The PDO index computed from HadISST (Figure 12a) shows a cold phase before the late 1970s and a transition to a warm phase afterward followed by a cold phase transition in the late 1990s. Compared with the UNINIT run (Figure 12b), the assimilated monthly PDO indices in both INIT_NOLOC and INIT runs (Figures 12c and 12d) correctly capture the two regime shifts with correlations of 0.47 and 0.50, respectively, much higher than that of the UNINIT run (−0.03). The 121-month running mean PDO index shows distinct decadal changes in all the experiments. The INIT_NOLOC and INIT runs exhibit more accurate decadal variabilities of the North Pacific SST, reaching correlations of 0.94 and 0.95, respectively, than the UNINIT run (0.41).

The dominant multidecadal SST oscillation in the North Atlantic is Atlantic Multidecadal Oscillation (AMO). The AMO index is defined as the linearly detrended SSTA averaged over the North Atlantic (0–60°N, //10.1029/2019MS001768 Journal of Advances in Modeling Earth Systems//).
The monthly AMO indices of all the experiments (Figures 13b–13d) show two regime shifts in the 1960s and 1990s, consistent with HadISST (Figure 13a). The correlation of the INIT_NOLOC run (0.49) is higher than that of the UNINIT run (0.42), although the amplitudes presented by both runs are underestimated. The 121-month running mean AMO indices of the UNINIT and INIT runs both correspond better with HadISST than the monthly versions. The good performance of the UNINIT run implies that the AMO is partly influenced by external forcings (Doblas-Reyes et al., 2013). The INIT run, where a localization is conducted to reduce the long-range spurious correlations, behaves better than the INIT_NOLOC run not only for the monthly but also the 121-month running mean AMO index.

The Atlantic meridional overturning circulation (AMOC) is a large-scale oceanic circulation from the mid to deep ocean (Delworth & Mann, 2000; Knight et al., 2005; Latif et al., 2004), which presents significant decadal to multidecadal variability. Following Karspeck et al., 2017, the multiyear mean and variability of AMOC are shown in Figure 14. The multiyear mean AMOC strength and structure of the UNINIT run (Figure 14a) exhibits net northward flow in the upper 1,000 m and southward flow below, which is similar to the No Assimilation forced experiments in Karspeck et al., 2017. The AMOC maximum, the depth of the southward return flow, and the deep ocean anticyclonic flow resemble some of the ocean reanalyses shown in Karspeck et al., 2017. The INIT_NOLOC run (Figure 14b) shows much stronger and deeper transport than the UNINIT run, which fails to agree with any ocean reanalysis in Karspeck et al., 2017. The standard deviation of the annual mean linearly detrended AMOC in the UNINIT run (Figure 14d) displays a similar performance to the No Assimilation set of forced model simulations shown in figure 5 of Karspeck et al., 2017. The INIT_NOLOC run (Figure 14e), with maxima in the tropical southern hemisphere and midlatitude northern hemisphere, presents a larger standard deviation than that in the UNINIT run. This pattern disagrees with any ocean reanalysis in Karspeck et al., 2017. The much stronger transport and variability of AMOC in the INIT_NOLOC run may result from the spurious correlations over long spatial distances in the vertical directions due to the small ensemble size (i.e., 30) in the ensemble-based initialization method (Evensen, 2009). Localization is one way to alleviate the spurious correlations (Evensen, 2009).
However, the INIT run, which includes the simplified off-line localization, shows little improvement (Figures 14c and 14f) comparing with the INIT_NOLOC run. Therefore, an online localization method is worth being tested in the future study.

4. Prediction Skill of the Hindcast Experiments

In this section, we preliminarily investigate the prediction skills of the annual mean SATA, PDO, and AMO indices. Both HCST_NOLOC and HCST runs are evaluated using the 10-member ensemble mean fields. The HCST run is compared with benchmarks of persistence, UNINIT, INIT, and HCST_NOLOC runs, focusing on the correlation with the verification data and RMSE as a function of the lead year. For each lead year, like lead year 1, time series of surface air temperature and SST from 1961 to 1997 are constructed from the first year of each ensemble mean hindcast. The same process is applied to lead year 2, generating time series from 1962 to 1998, until lead year 10 from 1970 to 2006. The time series of the verification data and the four benchmarks are computed the same way. All the experiments are bias corrected dependent on the lead time (ICPO, 2011). These time series are also linearly detrended to remove the linearly varying external forcing signals. The above time series are then used to calculate the prediction skills.

Correlations of global, ocean, and land mean SATA with HadCRUT4 are shown in Figure 15, which are generally higher than those after detrending, suggesting the role of external forcings (van Oldenborgh et al., 2012). An increasing trend of correlation is present for all the experiments except persistence along the lead year (Figures 15a–15c), implying a more critical role of external forcings at longer forecast lead time (Boer et al., 2013). The INIT run shows the highest correlation (Figure 15) and lowest RMSE (Figure S3 in supporting information) at all lead years for global and ocean mean SATA while at most lead years for detrended land mean SATA. The predicted land mean SATA even outperforms the INIT run with higher correlation (Figure 15c) and lower RMSE (Figure S3c in supporting information), which can be explained by both the external forcings and the transmission of ocean observational information to the atmosphere through air-sea interactions during the 1-month assimilation window. Compared to the UNINIT run, the HCST run has higher correlation at all lead years for global and land mean SATA while except for lead years 1–3 for ocean mean SATA; the RMSE of the HCST run is smaller at all lead years except for lead years 1–3 for detrended ocean mean SATA. Compared to persistence, the HCST run has higher correlations except for lead year 1 for ocean mean SATA and lead year 1 for detrended land mean SATA while lower RMSE for all cases. Compared to the HCST_NOLOC run, the HCST run are superior except for lead year 1 for ocean mean SATA and lead years 5, 7, and 9 for land mean SATA. The overall better performance of the HCST run than the HCST_NOLOC run mainly results from the localization to obtain more accurate ICs for decadal predictions. These results suggest that the high decadal prediction skills after detrending stem from the high-quality ICs obtained during the DRP-4DVar-based initialization while those with trend further incorporates the role of external forcings. Also, note that the skill of detrended ocean mean SATA in later lead years (9 and 10) is larger than in earlier lead years (2–5), which possibly results from the fact that more observational signal is released during the 10-year model hindcast integration through the model air-sea interactions. In addition, the coupled model tends to undergo adjustments to the ICs during the lead years 2–5, which probably degrades the prediction skills to some extent.

On smaller spatial scales, the North Atlantic subpolar gyre (SPG) and East Asia (Figures 16a and 16b and S4a and S4b in supporting information) are chosen as examples to investigate the prediction skills of linearly detrended SATA over these two regions. As for the detrended SATA over the North Atlantic SPG
Figure 13. Time series of monthly AMO index during the period of 1955–2006 from (a) HadISST, (b) UNINIT experiment, (c) INIT_NOLOC, and (d) INIT runs (unit: °C). The yellow (green) bar represents the warm (cold) phase of the AMO index. The solid blue line denotes 121-month running mean AMO indices with HadISST are given at the top left corner before and after the forward slash, respectively.

(Tables 16a and S4a in supporting information), the persistence shows the highest correlation for lead year 1, then sharply drops below the significance line until lead year 4 and increases till lead year 10 which passes the significant test. The UNINIT run fails to present significant correlation skills for all lead years. The RMSEs of persistence and the UNINIT run are lower than the other experiments. Significant correlations turn up in the INIT run for all lead times. The HCST run is superior to the INIT run for all lead years except for lead years 1 and 5 and the HCST_NOLOC run for about half of the lead years. The HCST run also performs better than persistence except for lead year 1 and the UNINIT run for all lead times. The RMSE of the HCST run is the smallest among the five experiments for most lead years. The higher skills of the HCST_NOLOC and HCST runs than persistence and the UNINIT run indicate that the prediction skill of North Pacific SPG relies mostly on the accurate ICs (i.e., the INIT run). The higher skills of the HCST_NOLOC and HCST runs than the INIT runs at most lead years suggest the role of air-sea interactions during model integrations for the hindcasts. The detrended SATA over East Asia (Figures 16b and S4b in supporting information), however, fails to show encouraging results as that over the North Atlantic SPG. The correlation skill of the INIT run are insignificant at all lead years, possibly due to the lack of atmospheric assimilation. The inaccurate ICs over East Asia thus lead to low prediction skills which are insignificant at most lead years. Although the prediction skill over East Asia is not good on the annual time scale, it improves substantially on the decadal time scale based on the detrended 10-year averaged SATA. The correlation of the detrended 10-year mean SATA in the HCST run is 0.64, which is higher than those in the HCST_NOLOC run (0.51), the UNINIT run (0.23), and persistence (0.22) and smaller than that in the INIT run (0.81).

The correlation and RMSE of the PDO and AMO indices are shown in Figures 16c and 16d and S4c and S4d in supporting information. For the PDO index (Figures 16c and S4c in supporting information), the persistence has the highest correlation for lead year 1 and then drops rapidly below the significance line. The INIT run has the highest correlation around 0.6–0.7 and the lowest RMSE at almost all lead years. The HCST run corresponds better with HadISST than the UNINIT run which is significantly correlated with HadISST and better than persistence for lead years 3–10. However, correlations of the HCST run during the first three lead years are not significant, which may result from the slowly released ocean signals in the ICs. The HCST run performs better than the HCST_NOLOC run for lead years 3–7 and 10. For the AMO index (Figures 16d and S4d in supporting information), the INIT run shows a correlation of around 0.5, a little smaller than that of the PDO index. The HCST run, with higher correlation and lower RMSE compared to those of the UNINIT run, outperforms the INIT run in several lead years. The HCST run is superior to the HCST_NOLOC run for about half of the lead years, while the correlation of the HCST_NOLOC run decreases for the last three lead years. We note that the correlations of the UNINIT run are statistically significant in the first eight lead years. These results indicate that the high prediction skill of the AMO index relies on both external forcings and ICs, which agrees with Doblas-Reyes et al. (2013).

5. Discussion and Conclusion

A DRP-4DVar-based CDA system with a simplified off-line localization technique was designed for a coupled model FGOALS-g2 and applied to decadal predictions. This system is one of the earliest 4DVar-based CDA systems. It consists of two steps. The first step is the DRP-4DVar assimilation without localization, which uses historical samples to estimate the background error covariance matrix and avoids the use of the adjoint model to save considerable computing costs. The second step is the simplified off-line localization, which is used to reduce the spurious long-distance correlations in the background.
error covariance matrix and save the large computing expense of localization. Then a 1-month free model integration is started from the analysis produced in the second step. It is the first time that the CDA system has been shown to have the ability to directly assimilate monthly mean observations without interpolating to each model time step, as is seen in nudging and IAU methods. It has a performance comparable to that of 4DVar but costs as less as a nudging-based IAU. Furthermore, the initialization only takes 2.375 times of the time cost of model integration in the assimilation window. However, the simplified off-line localization technique has several disadvantages. First, the principal modes used to generate the extended samples only takes account of one direction, leading to poor representativeness of the principal modes. Additionally, the analysis after localization cannot influence the background of the next month due to the off-line way of the localization. Thus, the current localization is insufficient to absorb the observed variability.

Figure 14. The multiyear mean AMOC stream function (top) and the standard deviation of the annual mean, linearly detrended AMOC stream function (bottom) from 1960 to 2006 for the UNINIT (left), INIT_NOLOC (middle) and INIT (right) runs. The bold line represents the zero and 0.5 Sv contour in (a)–(c) and (d)–(f), respectively.

Figure 15. Correlation of global (left), ocean (middle), and land (right) mean SATA with HadCRUT4 as a function of the lead year with (top) and without (bottom) the linear trend. The purple, green, blue, orange, and red lines denote persistence (PERS), UNINIT, INIT, HCST_NOLOC, and HCST experiments, respectively. The black line indicates the 95% confidence level of a two-sided t test.
A 62-year initialization experiment is conducted, and its performance is assessed in terms of bias, RMSE, climatological mean states, and climate variabilities. Compared with the UNINIT run, the biases of the 62-year averaged ocean temperature and salinity are smaller in the upper 600 m; the RMSEs are smaller at all depths, excluding temperature below 400 m and salinity between 100 and 500 m. The temporal evolutions of the assimilated SSTA and SSSA are closer to the verification data than those of the UNINIT run. The spatial distributions of the assimilated SST, SSS, and SSH averaged over 1955–2006 and precipitation averaged over 1979–2006 show reduced biases in some areas. The assimilated SATA averaged over the globe, ocean, and land corresponds better with the verification data than the UNINIT and INIT_NOLOC runs. As for climate variabilities in the Pacific and Atlantic, the phases of assimilated ENSO, PDO, and AMO are highly consistent with the verification data. However, the amplitudes of ENSO and AMO are slightly smaller than the verification data.

Starting with the ICs from the INIT run, 37 sets of 10-member ensemble hindcasts are carried out. The ensemble members are generated using a lagged initialization method. Other ensemble generation methods, such as Bred Vectors and Ensemble dispersion filter in Polkova et al. (2019), are not adopted in this study. Polkova et al. (2019) compared the decadal prediction skills using different ensemble generation methods and found that no single method exists that outperforms the others regarding all the metrics. Further experiments are called for to study the impact of different ensemble generation methods on the decadal prediction skills in our CDA system.

The HCST run shows that regardless of whether the trend is removed or not, the global, ocean, and land averages of SATA mostly shows higher prediction skills than those of persistence, the UNINIT run and the HCST_NOLOC run. The SATA prediction skill over the North Atlantic SPG in the HCST run is higher than persistence and the UNINIT run and comparable to the HCST_NOLOC run, while that over East Asia is high on the decadal time scale. The prediction skills of the PDO and AMO indices are higher than those of persistence and the UNINIT run at most lead times, although the correlations of the predicted PDO index in the first three lead years fail to be significant.

In the initialization experiment, the ocean assimilation conducted by the CDA system not only may directly constrain the ocean states but also may indirectly benefit the low-layer atmosphere in the regions where the ocean mainly forces the atmosphere (e.g., Tropic Pacific, many areas in North Atlantic), through the coupled model constraints during the coupled integrations in the 1-month assimilation window. The related
initialized decadal predictions may generally present higher skills at these regions. However, in the domains where the ocean is mainly forced by the atmosphere (e.g., many areas in North Pacific), the ocean assimilation may have little influence on the atmosphere. In particular, fewer ocean observations might be incorporated in these domains during the initialization because of biased atmospheric forcing provided by the model without constraints of atmospheric assimilation. This is why the corresponding initialized decadal predictions might perform worse in the domains. For this reason, consistent inclusion of atmospheric assimilation or even land and sea ice assimilations in the CDA system may be an important way to further increase the prediction skill over the regions where the ocean is mainly driven by other components.

In a word, the good performance of the initialization and reasonable prediction skill in the hindcasts suggest that the DRP-4DVar method is an advantageous initialization approach in the decadal prediction field. As the DRP-4DVar-based initialization is independent of the coupled model, more efforts will be made in its application to other model components (e.g., atmosphere, land, and sea ice), other coupled climate models or earth system models.

---

**References**

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., et al. (2003). The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydro meteorology*, 4(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPPC>2.0.CO;2

Alexander, M. A., Blace, L., 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), 2205–2231.

Balmaseda, M. A., Alves, O. J., Arruab, A., Awiij, T., Behringer, D. W., Perry, N., et al. (2009). Ocean initialization for seasonal forecasts. *Oceanography*, 22(3), 154–159.

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1132–1161.

Bellucci, A., Guidali, S., Masina, S., Storto, A., Scoccimarro, E., Cagnazzo, C., et al. (2013). Decadal climate predictions with a coupled OAGCM initialized with oceanic reanalyses. *Climate Dynamics*, 40(5), 1483–1497.

Boer, G., Kharin, V., & Merry, C. (2009). *Geoscientific Model Development*. 3(2), 154–120.

Balmaseda, M. A., Alves, O. J., Arribas, A., Awiij, T., Behringer, D. W., Perry, N., et al. (2009). Ocean initialization for seasonal forecasts. *Oceanography*, 22(3), 154–159.

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1132–1161.

Boer, G., Smith, D. M., Cassou, C., Doblas-Roy, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9(10), 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Carton, J. A., & Giese, B. S. (2008). A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Monthly Weather Review*, 136(3), 2999–3017.

Chen, D., Can, M., Kaplan, A., Zebiak, S. E., & Huang, D. (2004). Predictability of El Niño over the past 148 years. *Nature*, 428, 733.

Craig, A. P., Jacob, R. L., Kauffman, B., Bette, L., Larson, J., Ong, E., et al. (2005). Cpt6: The new extensible, high-performance parallel coupler for the community climate system model. *International Journal of High Performance Computing Applications*, 19(3), 309–317.

Crawford, W., Martinez, R., & Suga, T. (2006). List of possible applications for decadal prediction, CLIVAR Pacific Panel Document available at https://www.clivar.org/organization/pacific/DecadalPredictions.pdf.

Delworth, T. L., & Mann, M. E. J. (2000). Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dynamics*, 16(9), 661–676.

Doblas-Reyes, F. J., Andrade-Llorio, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., et al. (2013). Initialized near-term regional climate change predictions. *Nature Communications*, 4, 1713.

Du, H., Doblas-Reyes, F. J., García-Serrano, J., Guemas, V., Soufflet, Y., & Wouters, B. (2012). Sensitivity of decadal predictions to the initial atmospheric and oceanic perturbations. *Climate Dynamics*, 39(7), 2013–2023.

Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter*. Dordrecht, Heidelberg: Springer Science & Business Media. https://doi.org/10.1007/978-3-642-04311-5

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(5), 1937–1958.

Frankignoul, C., & Hasselmann, K. (1977). Stochastic climate models, part II application to sea-surface temperature anomalies and thermocline variability. *Tellus*, 29(4), 289–305.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723–737.

Gill, A. (1980). Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, 106(449), 447–462.

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113, D06104. https://doi.org/10.1029/2007JD008972

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(1–2), 245–272.

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118, 6704–6716. https://doi.org/10.1002/2013JC009067

Ham, Y.-G., Rienecker, M. M., Suarez, M. J., Vihlikaev, Y., Zhao, B., Marshak, J., et al. (2014). Decadal prediction skill in the GEOF-S forecast system. *Climate Dynamics*, 42(1), 1–20.

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776–2790.

He, Y., Wang, B., Liu, M., Liu, L., Yu, Y., Liu, J., et al. (2017). Reduction of initial shock in decadal predictions using a new initialization strategy. *Geophysical Research Letters*, 44, 5838–5847. https://doi.org/10.1002/2017GL074028

IPCC (International CLIVAR Project Office) (2011). Data and bias correction for decadal climate predictions. *International CLIVAR Project Office*, CLIVAR Publication Series, 150, 6.
Ishii, M., Kimoto, M., Sakamoto, K., & Iwaski, S.-I. (2006). Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. Journal of Oceanography, 62(2), 155–170.

Ishii, M., Shouji, A., Sugimoto, S., & Matsumoto, T. (2005). Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. International Journal of Climatology, 25(7), 865–879.

Kalnay, E. (2003). Atmospheric Modeling, Data Assimilation and Predictability. New York: Cambridge university press.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP-NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society, 77(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Karspeck, A., Yeager, S., Danabasoglu, G., & Teng, H. (2015). An evaluation of experimental decadal predictions using CCSM4. Climate Dynamics, 44(3), 907–923.

Karspeck, A. R., Stammer, D., Kohl, A., Danabasoglu, G., Balmaseda, M., Smith, D. M., et al. (2017). Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. Climate Dynamics, 49(3), 957–982. https://doi.org/10.1007/s00382-015-2787-7

Keenlyside, N., Latif, M., Botzet, M., Jungclaus, J., & Schulzweida, U. (2005). A coupled model for initializing El Nino Southern Oscillation forecasts using sea surface temperature. Tellus A, 57(3), 340–356.

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornbluh, L., & Roeckner, E. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector. Nature, 453, 84.

Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M., & Mann, M. E. J. (2005). A signature of persistent natural thermohaline circulation cycles in observed climate. Geophysical Research Letters, 32, L20708. https://doi.org/10.1029/2005GL024233

Latif, M., Roeckner, E., Botzet, M., Esch, M., Haak, H., Hagemann, S., et al. (2004). Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic Thermohaline circulation with sea surface temperature. Journal of Climate, 17(7), 1605–1614. https://doi.org/10.1175/1520-0442(2004)017<1605:RMMDTC>2.0.CO;2

Li, L., Lin, P., Yu, Y., Wang, B., Zhou, T., Liu, L., et al. (2013). The Flexible Global Ocean–Atmosphere–Land System Model. Grid-point Version 2: FGOSLs-g2. Advances in Atmospheric Sciences, 30(3), 543–560. https://doi.org/10.1007/s00376-012-2140-6

Li, L., Wang, B., Dong, L., Liu, L., Shen, S., Hu, N., et al. (2013). Evaluation of Grid-Point Atmospheric Model of IAP LASG version 2 (GAMIL2). Advances in Atmospheric Sciences, 30(3), 855–867. https://doi.org/10.1007/s00376-013-2157-5

Lindzen, R. S., & Nigam, S. (1987). On the role of sea surface temperature gradients in forcing low-tropics.

Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. Quarterly Journal of the Royal Meteorological Society, 129(595), 3183–3203.

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(6), 1099–1108.

McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2003). ENSO as an integrating concept in earth science. Science, 314(5806), 1740–1745. https://doi.org/10.1126/science.1132588

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction. Bulletin of the American Meteorological Society, 90(10), 1467–1486. https://doi.org/10.1175/2009BAMS2778.1

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian seasonal to interannual prediction system. Part I: Models and Initialization. Monthly Weather Review, 141(8), 2910–2945.

Mochizuki, T., Masuda, S., Ishikawa, Y., & Awoji, T. (2016). Multiyear climate prediction with initialization based on 4D-Var data assimilation. Geophysical Research Letters, 43, 3903–3910. https://doi.org/10.1002/2016GL067895

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. Journal of Geophysical Research, 117, D08101. https://doi.org/10.1029/2011JD017187

Oke, P. R., Brassington, G. B., Griffin, D. A., & Schiller, A. (2010). Ocean data assimilation: a case for ensemble optimal interpolation. Australian Meteorological and Oceanographic Journal, 59, 67–76.

Oke, P. R., Sakov, P., & Corny, S. P. (2007). Impacts of localisation in the EnKF and EnOI: Experiments with a small model. Ocean Dynamics, 57(1), 32–45.

Oleson, K. W., et al. (2010). Technical description of version 4.0 of the Community Land Model (CLM), NCAR Technical Note NCAR/TN-478+STR, National Center for Atmospheric Research.

Pan, Y. H., & Oort, A. H. (1983). Global climate variations connected with sea surface temperature anomalies in the eastern equatorial Pacific Ocean for the 1958–73 period. Monthly Weather Review, 111(6), 1244–1258.

Polkova, I., Brune, S., Kadow, C., Romanova, V., Gollan, G., Baehr, J., et al. (2019). Initialization and ensemble generation for decadal climate predictions: A comparison of different methods. Journal of Advances in Modeling Earth Systems, 11(1), 149–172. https://doi.org/10.1029/2018MS001439

Rayner, N., Parker, D. E., Horton, E., Folland, C., Alexander, L., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research, 108(D14), 4047. https://doi.org/10.1029/2002JD002670

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8), 1055–1058. https://doi.org/10.1175/2010BAMS0011.1

Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., & Terray, L. (2016). Drift dynamics in a coupled model initialized for decadal forecasts. Climate Dynamics, 46(5), 1819–1840.

Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. Climate Dynamics, 41(11), 3325–3338.

Sugiura, N., Awoji, T., Masuda, S., Mochizuki, T., Tomyoda, T., Miyama, T., et al. (2008). Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. Journal of Geophysical Research, 113, C10017. https://doi.org/10.1029/2008JC004714

Swingedouw, D., Mignot, J., Labetouille, S., Guilyardi, E., & Madec, G. (2013). Initialisation and predictability of the AMOC over the last 50 years in a climate model. Climate Dynamics, 40(9), 2381–2399.

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. Ser. II, 90A, 275–294. https://doi.org/10.2151/jmsj.2012-A14
Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research, 106(D7), 7183–7192.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485–498.

Trenberth, K. E. (1984). Signal versus noise in the Southern Oscillation. Monthly Weather Review, 112(2), 326–332.

Trenberth, K. E., & Shea, D. J. (2006). Atlantic hurricanes and natural variability in 2005. Geophysical Research Letters, 33, L12704. https://doi.org/10.1029/2006GL026894

Troccoli, A., & Palmer, T. N. (2007). Ensemble decadal predictions from analysed initial conditions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2179–2191.

van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., & Hazeleger, W. (2012). Decadal prediction skill in a multi-model ensemble. Climate Dynamics, 38(7), 1263–1280.

Wang, B., Liu, J., Liu, L., Xu, S., & Huang, W. (2018). An approach to localization for ensemble-based data assimilation. PLoS ONE, 13(1), e0191088. https://doi.org/10.1371/journal.pone.0191088

Wang, B., Liu, J., Wang, S., Cheng, W., Juan, L., Liu, C., et al. (2010). An economical approach to four-dimensional variational data assimilation. Advances in Atmospheric Sciences, 27(4), 715–727.

Wang, B., Liu, M., Yu, Y., Li, L., Lin, P., Dong, L., et al. (2013). Preliminary evaluations of FGOALS-g2 for decadal predictions. Advances in Atmospheric Sciences, 30(3), 674–683. https://doi.org/10.1007/s00376-012-2084-x

Wei, M., Li, Q., Xin, X., Zhou, W., Han, Z., Luo, Y., & Zhao, Z. (2017). Improved decadal climate prediction in the North Atlantic using EnOI-assimilated initial condition. Science Bulletin, 62(16), 1142–1147.

Wu, B., Chen, X., Song, F., Sun, Y., & Zhou, T. (2015). Initialized decadal predictions by LASG/IAP climate system model FGOALS-s2: Evaluations of strengths and weaknesses. Advances in Meteorology, 2015, 12.

Wu, B., Zhou, T., & Zheng, F. (2018). EnOI-IAU initialization scheme designed for decadal climate prediction system IAP-DecPreS. Journal of Advances in Modeling Earth Systems, 10(2), 342–356.

Xin, X., Wu, T., & Zhang, J. (2013). Introduction of CMIP5 experiments carried out with the climate system models of Beijing Climate Center. Advances in Climate Change Research, 4(1), 41–49.

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. Journal of Climate, 25(15), 5173–5189.

Yu, Y., Zhi, H., Wang, B., Wan, H., Li, C., Liu, H., et al. (2008). Coupled model simulations of climate changes in the 20th century and beyond. Advances in Atmospheric Sciences, 25(4), 641–654.

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Monthly Weather Review, 135(10), 3541–3564.

Zhang, W., Lin, J., Xu, W., Fu, H., & Yang, G. (2017). SCStore: Managing scientific computing packages for hybrid system with containers. Tsinghua Science and Technology, 22(6), 675–681.