Revisiting online and offline data assimilation comparison for paleoclimate reconstruction: an idealized OSSE study

Okazaki Atsushi[1], Miyoshi Takemasa[2], Yoshimura Kei[3], Greybush Steven J.[1], and Zhang Fuqing[1]

[1]Pennsylvania State University
[2]RIKEN
[3]University of Tokyo

November 16, 2022

Abstract

Data assimilation (DA) has been applied to estimate the time-mean state such as annual mean surface temperature for paleoclimate reconstruction. There are two types of DA for this purpose: online-DA and offline-DA. The online-DA estimates the time-mean states and the initial conditions for the next DA cycles while the offline-DA only estimates the former. If there is sufficiently long predictability in the system of interest compared to the temporal resolution of the observations, online-DA is expected to outperform offline-DA by utilizing information in the initial conditions. However, previous studies failed to show the superiority of online-DA when time-averaged observations are assimilated, and the reason has not been investigated thoroughly. This study compares online-DA and offline-DA and investigates the relation to the predictability using an intermediate complexity general circulation model with perfect-model observing system simulation experiments. The result shows that the online-DA outperforms offline-DA when the length of predictability is longer than the averaging time of the observations. We also found that the longer the predictability, the more skillful the online-DA. Here, the ocean plays a crucial role in extending predictability, which helps online-DA to outperform offline-DA. Interestingly, the observations of near-surface air temperature over land are found to be highly valuable to update the ocean variables in the analysis steps, suggesting the importance to use cross-domain covariance information between the atmosphere and the ocean when online-DA is applied to reconstruct paleoclimate.

Posted on 16 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10504749.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Revisiting online and offline data assimilation comparison for paleoclimate reconstruction: an idealized OSSE study

A. Okazaki[1,2,3], T. Miyoshi[2,4,5,6,7,8], K. Yoshimura[3], S. J. Greybush[1], and F. Zhang[1]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA

[2]RIKEN Center for Computational Science, Kobe, Hyogo, Japan

[3]Institute of Industrial Science, The University of Tokyo, Kashiwa, Chiba, Japan

[4]RIKEN interdisciplinary Theoretical and Mathematical Sciences Program, Kobe, Hyogo, Japan

[5]RIKEN Cluster for Pioneering Research, Kobe, Hyogo, Japan

[6]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, MD, USA

[7]Graduate School of Science, Kyoto University, Kyoto, Kyoto, Japan

[8]Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan

Corresponding author: Atsushi Okazaki (ats.okazaki@gmail.com)

Key Points:

- Online-DA outperforms offline-DA if there is practical predictability longer than the temporal resolution of the observations
- The ocean plays a vital role for online-DA to outperform offline-DA when assimilating time-mean observations
- Observations over land are valuable to update the ocean, suggesting strongly coupled-DA should be used for paleoclimate reconstruction

Abstract

Data assimilation (DA) has been applied to estimate the time-mean state such as annual mean surface temperature for paleoclimate reconstruction. There are two types of DA for this purpose: online-DA and offline-DA. The online-DA estimates the time-mean states and the initial conditions for the next DA cycles while the offline-DA only estimates the former. If there is sufficiently long predictability in the system of interest compared to the temporal resolution of the observations, online-DA is expected to outperform offline-DA by utilizing information in the initial conditions. However, previous studies failed to show the superiority of online-DA when time-averaged observations are assimilated, and the reason has not been investigated thoroughly. This study compares online-DA and offline-DA and investigates the relation to the predictability using an intermediate complexity general circulation model with perfect-model observing system simulation experiments. The result shows that the online-DA outperforms offline-DA when the length of predictability is longer than the averaging time of the observations. We also found that the longer the predictability, the more skillful the online-DA. Here, the ocean plays a crucial role in extending predictability, which helps online-DA to outperform offline-DA. Interestingly, the observations of near-surface air temperature over land are found to be highly valuable to update the ocean variables in the analysis steps, suggesting the importance to use cross-domain covariance information between the atmosphere and the ocean when online-DA is applied to reconstruct paleoclimate.

1 Introduction

Knowledge of past climate conditions is crucial to understanding the climate system. Recently, data assimilation (DA) has been applied to reconstruct paleoclimate of the last millennium (e.g. Franke et al., 2017; Goosse et al., 2010; 2012; Hakim et al., 2016; Steiger et al., 2018; Tardif et al., 2019) and ages deeper into the past (e.g. Kurahashi-Nakamura et al., 2017; Mathiot et al., 2013; Renssen et al., 2015). DA estimates the most likely state (analysis) by combining model simulations (background) and observations. DA has long been used for numerical weather prediction (NWP) and is a well-established method (e.g. Kalnay, 2003; Houtekamer and Zhang, 2016; and references therein). The purpose of DA in NWP is to provide the optimal initial conditions for subsequent forecasts, where the analyses represent instantaneous values and will be used as initial conditions for the next DA cycles. On the other hand, DA has been used differently in the application to paleoclimate reconstruction, where DA is used to estimate time-mean states. Here, for the paleoclimate reconstruction, climate proxies such as physical and chemical properties of ice cores, corals, trees, speleothems, etc., which generally represent surface variables (e.g. surface temperature), serve as the observations. Those data represent time-averaged values (typically annual or longer) while the observations for NWP represent instantaneous values. To assimilate such time-averaged observations effectively, several DA methods have been proposed. Although there are a few methods based on nudging (von Storch et al., 2000), forcing singular vectors (van der Schrier and Barkmeijer, 2005), and the variational method (Kurahashi-Nakamura et al., 2017), this study focuses on ensemble-based ones. The first study that proposes ensemble-based method for paleoclimate reconstruction is Dirren and Hakim (2005): they proposed an Ensemble Kalman Filter (EnKF) based method that updates only the time-averaged component of state variables. Subsequently, Annan and Hargreaves (2012) and Goosse et al. (2012a, 2012b) applied a particle filter (PF; van Leeuwen, 2009). Those methods estimate not only time-mean states but also initial conditions for the next DA cycles as in NWP

and are categorized as online-DA methods. Huntley and Hakim (2010) showed the capability of online-DA when assimilating time-averaged data using quasi-geostrophic model (QG-model; Hakim, 2000). They also showed that the skill of online-DA decreases with increasing observation averaging time, and the advantage of the online-DA vanishes and reduces to that of non-initialized simulation when assimilating observations whose averaging time is longer than the predictability of the system. Pendergrass et al. (2012) used a QG-model coupled with a slab ocean and showed that the skill of the online-DA increases as ocean depth increases.

There is another method to estimate time-averaged states, known as "offline-DA" (Goosse et al., 2006; Bhend et al., 2012; Steiger et al., 2014). In offline-DA, only time-averaged states are analyzed but the initial conditions for the next DA-cycles are not. Hence, the forecasts are nothing more than free runs. The idea behind the method is that the observations are temporally too sparse to constrain the model. As mentioned above, most typically the observations in paleoclimate era represent an annual mean or longer. In other words, they are available once in a year at most. On the other hand, the predictability of the atmosphere is roughly 2 weeks (e.g. Lorenz, 1963; 1969). Therefore, the information contained in initial conditions will be lost long before the end of each DA cycle no matter how the initial conditions are close to the truth. Therefore, at least for the studies using only atmospheric general circulation models (AGCMs), there is no point to analyze initial conditions for the next DA cycles. There are two types of offline-DA; transient offline-DA and stationary offline-DA. The former's background consists of an ensemble run for the same time as observations represent (Bhend et al., 2012; Franke et al., 2017; Goosse et al., 2006). On the other hand, the latter's background consists of a single run and is the same at all the analysis steps (Hakim et al., 2016; Steiger et al., 2014; Tardif et al., 2019). The idea of stationary offline-DA is similar to that of ensemble optimal interpolation (EnOI; Oke et al., 2002; 2005). The stationary offline-DA is computationally efficient as it does not require an ensemble run. Steiger et al. (2014) showed that the method outperforms the conventional climate field reconstruction method based on principal component analysis in reconstructing the surface temperature field, while its relative performance to transient offline-DA is still an open question.

Which DA method is favorable for paleoclimate reconstruction, online-DA or offline-DA? In terms of skill in estimating time-mean states, online-DA should be better than offline-DA if the system of interest has long enough predictability compared to the averaging time that observations represent since online-DA can use better initial conditions and flow dependent covariances in the case of EnKF-based methods. In terms of computational cost, offline-DA is preferable as it can use pre-existing model simulations and ensemble simulations are not prerequisite for stationary offline-DA, while these are not the case with online-DA. To efficiently use information contained in initial conditions with online-DA, models that contain slowly changing components, such as atmosphere-ocean coupled general circulation models (CGCMs) must be used. However, ensemble-based online-DA with a CGCM is prohibitively expensive. Due to this limitation, online-DA has been applied together with intermediate complexity models to assimilate decadal mean data (e.g. Goosse et al., 2010; 2012) or with a stochastic model to assimilate annual mean data (Perkins et al., 2017), and few studies have applied online-DA with a CGCM to assimilate annual mean observations.

Several studies compared the skill of online-DA to that of offline-DA. Matsikaris et al. (2015) used a CGCM known as Max Planck Institute for Meteorology Earth System Model (MPI-ESM), to assimilate continental scale decadal mean data with an ensemble-based method proposed by

Goosse et al. (2006) that selects a member that fits the observations best. Acevedo et al. (2017) used an intermediate complexity AGCM known as SPEEDY (Molteni, 2003; Kucharski et al., 2006) coupled with a slab ocean model to assimilate annual mean data with the EnKF-based method proposed by Dirren and Hakim (2005). Neither of the studies, however, showed superiority of online-DA over offline-DA in spite of the fact that both studies included a slowly changing component. They suspect this may be because the predictability is not long enough, or the initial conditions are poorly constrained by DA. Owing to the comparable skills, computational efficiency, and less complexity, nowadays the offline-DA has been widely used for paleoclimate applications (e.g. Dee et al., 2017; Franke et al., 2017; Okazaki and Yoshimura, 2017; Steiger et al., 2018; Tardif et al., 2014; 2015; 2019). However, considering that plenty of studies show that the predictability of the surface temperatures, which most proxies represent, are longer than annual (e.g. Collins et al., 2002; Doblas-Reyes et al., 2013), online-DA must be better than offline-DA when models that have slowly changing components, such as CGCMs, are used. If so, why did the previous studies fail to show the superiority of online-DA? Is this because of the lack of the predictability in the models used, and/or a poor representation of the initial condition?

This study investigates the capability of online-DA relative to offline-DA together with its relevance to predictability using the SPEEDY coupled with a slab ocean model within an idealized framework for the last millennium climate reconstruction, where relatively large number of proxies whose resolution is shorter than annual are available. The SPEEDY has an ability to represent realistic physics with low computational cost, which makes it perfectly suited for the feasibility study. It should be worthwhile to investigate this issue with ensemble-based online-DA with a CGCM being a feasible choice in the near future as computation power steadily increases. We also investigate for what length of averaging period online-DA outperforms offline-DA for time-averaged observations (e.g. paleoclimate observations). Since monthly mean observations are available (e.g. for stable water isotope ratios in corals as in Grottoli, 2006), it is also worthwhile to answer such a question. We will discuss implications to the real case based on the results of the idealized experiments.

This study involves a CGCM, data assimilation, and hence atmosphere-ocean coupled-DA (CDA). There are two types of CDA; weakly coupled-DA (WCDA) and strongly coupled-DA (SCDA) (Penny and Hamill, 2017 and references therein). WCDA uses a coupled model to generate the background but updates the analysis separately for each component. Namely, atmospheric (oceanic) observations are used to update only atmospheric (oceanic) state. On the other hand, atmospheric (oceanic) observations are used to update both atmosphere and ocean states using the intra-domain and cross-domain error covariance in SCDA. SCDA approach has desirable aspects compared to WCDA; it can use a larger number of observations to update analysis, which not only reduces uncertainty in the analysis but also achieves more self-consistent analysis by reducing the initialization shocks that can be caused by WCDA. Indeed, several studies showed superiority of SCDA (e.g. Sluka et al., 2016). However, SCDA does not always produce better results (e.g. Han et al., 2013) and a valid method for SCDA has yet to be established. With a series of CDA experiments conducted in this study, this study may also have implications beyond paleoclimate DA, to CDA and prediction on a variety of time scales.

This paper is structured as follows. Section 2 describes the method to assimilate time-averaged observations and to measure predictability. Section 3 presents the results and a discussion. The summary follows in Section 4.

2 Methods

2.1 Model

The SPEEDY model is an intermediate complexity AGCM (Molteni, 2003; Kucharski et al., 2006). SPEEDY is a hydrostatic model with a primitive equation dynamics and simplified parameterizations that include essential processes such as convection, large-scale condensation, longwave and shortwave radiation, and boundary layer turbulences. SPEEDY has been used to study atmospheric predictability (e.g. Abid et al., 2015; Bahaga et al., 2015; Ehsan et al., 2013) and has been used as a testbed for DA studies first by Miyoshi (2005) and a number of follow-on studies (e.g. Amezucua et al., 2014; Greybush et al., 2011; Hatfield et al., 2018; Kalnay et al. 2007; Kondo et al., 2016; 2019; Kotsuki et al., 2020; Li et al., 2009; Miyoshi et al., 2011; 2014). We choose SPEEDY for its ability to represent realistic physics with a low computational cost. Though SPEEDY does not consider a diurnal cycle, which makes it difficult to accurately represent convection and sub-grid-scale vertical heat transport (Molteni, 2003), this will not be a major issue for our study with an OSSE flamework. SPEEDY has a horizontal resolution of T30 (approximately 400 km on the equator) and 8 vertical levels. As an optional facility, SPEEDY can be coupled with a slab ocean model, which does not consider advection or convection but exchanges heat with the atmosphere at the ocean surface. The prognostic variable of the slab ocean model is only sea surface temperature (SST), and the differential equation is given by

$$\Delta SST_{t+1} = \frac{\tau_{ocn}}{\tau_{ocn}+\delta t}\Delta SST_t + \frac{\delta t}{d_{ocn}C_{ocn}}\Delta F_t \quad (1)$$

where ΔSST_t and ΔF_t are the SST and the net heat flux anomaly from the climatological mean at time t, τ_{OCN} the damping timescale (90 d), δt the timestep (1 d), d_{OCN} the depth of the slab ocean (50 m), and C_{OCN} the heat capacity (4.18×10^6 $JK^{-1}m^{-1}$) (http://users.ictp.it/~kucharsk/speedy_description/km_ver40_appendixA.pdf). The net heat flux is given by a sum of shortwave and longwave radiation, and sensible heat and moisture fluxes. Here, the sensible heat flux is a function of SST and near-surface air temperature, and the moisture flux is a function of SST and near-surface specific humidity. Through the heat exchange, the ocean affects the evolution of the atmosphere. Since the ocean has a large heat capacity and hence slow variability, it serves as the source of a long-term memory for the atmosphere. The length of the ocean memory can be tuned by changing τ_{ocn} and/or d_{ocn}. A simple illustrative example would be the case with infinitely large τ_{ocn} and d_{ocn}; in this example, the coefficients of ΔSST_t and ΔF_t in Eq. 1 converge to one and zero, respectively, bringing the system to steady state where the initial condition persists forever ($\Delta SST_{t+1} = \Delta SST_t = \Delta SST_{t-1} = \cdots = \Delta SST_{t=0}$). Therefore, the system has eternal predictability. We choose to couple SPEEDY with the slab ocean model to elongate the predictability of the system so that it can effectively assimilate time-averaged observations whose averaging time is beyond the predictability of the atmosphere and to investigate the response of online-DA and offline-DA skills to the length of predictability. The configuration of SPEEDY is the same as Acevedo et al. (2017).

2.2 Data assimilation method

This study uses the time-average update (TAU) method proposed by Dirren and Hakim (2005). The method updates only the time-averaged component of the model state and leaves the deviation from the time-mean unchanged. The time-averaged and the deviation represent the low- and high-frequency component of the system, respectively. Let $\mathbf{x}^b_{t_n}$ be an *N*-dimensional background model forecast at time t_n. The TAU first decomposes $\mathbf{x}^b_{t_n}$ into the time-averaged component $\overline{\mathbf{x}^b}_{t_{n-1:}t_n}$ and the deviation from it $\mathbf{x}^b_{t_n}{}'$:

$$\mathbf{x}^b_{t_n} = \overline{\mathbf{x}^b}_{t_{n-1:}t_n} + \mathbf{x}^b_{t_n}{}' \tag{2}$$

$$\overline{\mathbf{x}^b}_{t_{n-1:}t_n} = \frac{1}{\tau}\sum_{t=t_{n-1}}^{t_n} \mathbf{x}^b_t \tag{3}$$

where, $\tau = t_n - t_{n-1}$ defines the time-averaging length. For the time being, the time notations will be omitted for simplicity. After the decomposition, the TAU updates the time-averaged component. This study uses the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007), a variant of the EnKF (Evensen, 1994), to update the time-averaged component. Here, we consider an ensemble of *m* members. In LETKF, the analysis ensemble mean and the ensemble perturbations are given by:

$$\langle\overline{\mathbf{x}^a}\rangle = \langle\overline{\mathbf{x}^b}\rangle + \overline{\mathbf{X}^b}\widetilde{\mathbf{P}^a}\left(\mathbf{H}\overline{\mathbf{X}^b}\right)^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathcal{H}(\langle\overline{\mathbf{x}^b}\rangle)) \tag{4}$$

$$\overline{\mathbf{X}^a} = \overline{\mathbf{X}^b}\sqrt{m-1}\widetilde{\mathbf{P}^a}^{1/2} \tag{5}$$

where, $\langle\overline{\mathbf{x}^{a,b}}\rangle$ denotes the ensemble mean and $\overline{\mathbf{X}^{a,b}}$ the ensemble perturbation matrix, and the superscripts a and b denote analysis and background, respectively. Here, $\overline{\mathbf{X}^{a,b}}$ is the $N \times m$ matrix whose *i* th column is $\overline{\mathbf{x}^{a,b\,(i)}} - \langle\overline{\mathbf{x}^{a,b}}\rangle$, where superscript (i) denotes that it is the *i* th member of the ensemble ($i = \{1,2,\ldots,m\}$). The notation $\mathbf{y}^o$ is the observation vector of length *p*, $\mathbf{R}$ the observation error covariance matrix whose size is $p \times p$, $\mathcal{H}$ the observation operator that converts the model state to the observation equivalent quantity, $\mathbf{H}$ the linearized observation operator whose size is $p \times N$, and $\widetilde{\mathbf{P}^a}$ the covariance matrix in the ensemble space given by

$$\widetilde{\mathbf{P}^a} = \left[(m-1)\mathbf{I} + \left(\mathbf{H}\overline{\mathbf{X}^b}\right)^T\mathbf{R}^{-1}\left(\mathbf{H}\overline{\mathbf{X}^b}\right)\right]^{-1}. \tag{6}$$

Then the *i* th member of time-averaged analysis $\overline{\mathbf{x}^a}$ is obtained by adding $\langle\overline{\mathbf{x}^a}\rangle$ to the *i* th column of $\overline{\mathbf{X}^a}$. Finally, the time-averaged analysis and the background anomaly are combined to obtain the full analysis field at time t_n:

$$\mathbf{x}^a_{t_n} = \overline{\mathbf{x}^a}_{t_{n-1:}t_n} + \mathbf{x}^b_{t_n}{}'. \tag{7}$$

In online-DA, $\mathbf{x}^a_{t_n}$ is used as the initial condition for the next forecast to obtain $\mathbf{x}^b_{t_n+1}$, while $\mathbf{x}^b_{t_n}$ is used in the transient offline-DA. In stationary offline-DA, model integration is not involved, and $\mathbf{x}^b$ is the same at all the time steps ($\mathbf{x}^b_{t_n} = \mathbf{x}^b_{t_{n-1}} = \cdots = \mathbf{x}^b_{t_1}$).

2.3 Experimental design

We perform a series of observing system simulation experiments (OSSEs) with the perfect model assumption where model errors are excluded. The nature run (i.e. the truth) is created by

integrating SPEEDY for 100 years from an arbitrary date denoted Jan 1st 1900 after a 400-year spin-up period, where the model starts from the atmosphere at rest (wind speed is zero everywhere). The nature run is externally forced by solar and greenhouse gasses, but the irradiance and the concentration are both constant over time.

We created seven sets of the observation. Each set of the observation differs in the length of averaging time (τ). The values of τ considered are 1 d, 3 d, 10 d, 1 month, 3 months, 6 months, and 12 months. Surface temperature data are drawn from the nature run, and the observations are generated in a way that mimics the real proxy network for the last millennium that consists of the composite of Mann et al. (2008), PAGES 2k consortium (2013), and Okazaki and Yoshimura (2017). When an observation is on the ocean grid in SPEEDY, it is assumed to represent sea surface temperature (SST). Otherwise, the observations represent surface air temperature (SAT). Therefore, SAT (SST) observations are all over land (ocean). This results in 413 SAT and 34 SST observation points (Figure 1(a)). Each observation is assumed to have an error of 1/2 of the climatological variance for averaging time of the observations (Figures 1(b)-(h)). Note that the seasonality is subtracted before calculating the variance when the averaging time is shorter than 12 months.

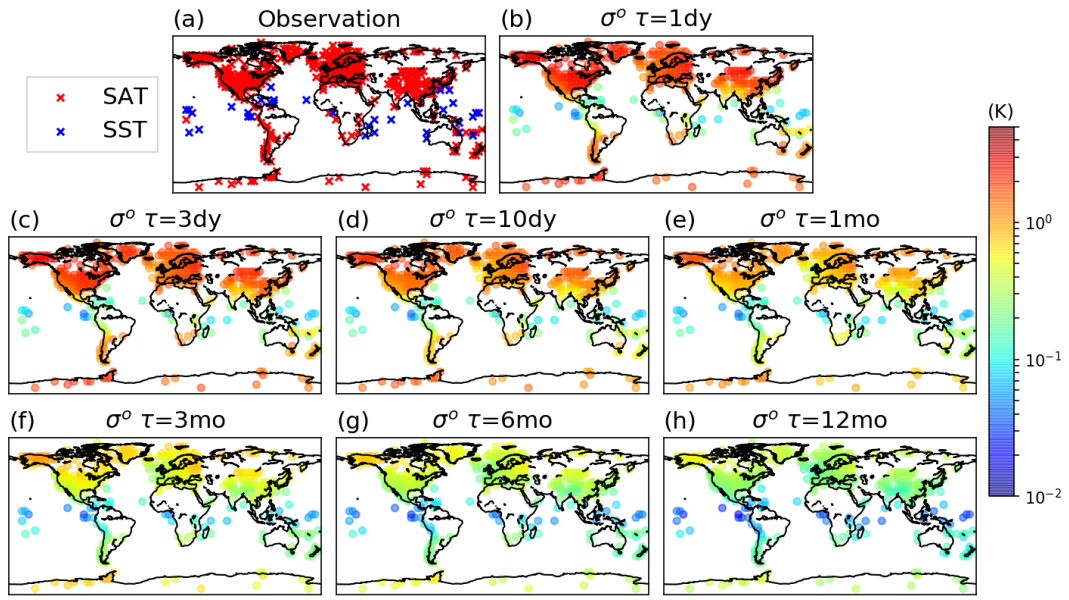

Figure 1 (a) Observation network and observation error standard deviations for τ of (b) 1 d, (c) 3 d, (d) 10 d, (e) 1 month, (f) 3 months, (g) 6 months, and (h) 12 months. Red and blue crosses in (a) denotes SAT and SST observation points, respectively. The unit of the errors is Kelvin.

Data assimilation experiments are performed with online-DA (ONLINE) and offline-DA (OFFLINE), and without DA (NODA) (Table. 1). For experiments with offline-DA, transient offline-DA is used in this study. Each experiment consists of seven subexperiments, and each

subexperiment except for NODA assimilates one of the observation sets. That is, the subexperiment 1, 2, ..., 7 assimilates observations with τ of 1 d, 3 d, 10 d, 1 month, 3 months, 6 months, and 12 months, respectively. Note that the analysis interval of DA-cycles is the same as the averaging time τ of the observation. Namely, the analysis interval of DA-cycle is different in each experiment. Each experiment consists of one ensemble integration of 100 DA-cycles starting from Jan 1st 1900. For all the experiments, the ensemble size is fixed to 30. We determined the ensemble size by considering that at this time the computation is too costly to run a larger ensemble with a full CGCM when online-DA is applied to the real problem. For both ONLINE and OFFLINE, the atmospheric (oceanic) observations are used to update both atmosphere and ocean states using the intra-domain and cross-domain error covariance (i.e. SCDA). The initial conditions for all the experiments are identical and are drawn from dates in subsequent years of the nature run (i.e. Jan 1st 2000, 2001,…, 2029).

To investigate response of skills in online-DA and offline-DA to the length of predictability, another set of experiments similar to the above but with τ_{OCN} of 360 d are performed (Table. 1). After changing the parameter value, we re-generated the nature run and the observations in the same way and repeated all the DA experiments. The experiments with τ_{ocn} of 360 d will be referred to as ONLINE_LONG, OFFLINE_LONG, and NODA_LONG. Note that the initial conditions used in the experiments are different from those used in ONLINE, OFFLINE, and NODA as they are drawn from the respective nature run.

To investigate the relative impact of the atmosphere and ocean in ONLINE, another four experiments with online-DA are performed (Table. 1); ONLINE_AA, ONLINE_OO, ONLINE_AO, and ONLINE_OA. The ONLINE_AA (ONLINE_OO) assimilates only SAT (SST) to update atmospheric (oceanic) variables, and the oceanic (atmospheric) variables are not updated. Namely, only intra-domain information is used but cross-domain information is not in ONLINE_AA and ONLINE_OO in the analysis steps. Conversely, only cross-domain information is used in ONLINE_AO and ONLINE_OA: the ONLINE_AO (ONLINE_OA) assimilates only SAT (SST) to update oceanic (atmospheric) variables. The improvement in SST in ONLINE_AA and ONLINE_OA compared to NODA is due to the interaction between the atmosphere and ocean in the forecast steps. On the other hand, the improvement in SST in ONLINE_AO and ONLINE_OO is due to updating the oceanic variables in analysis steps. Note that the initial conditions used in the experiments are identical with those used in ONLINE, OFFLINE, and NODA.

For experiments using online-DA, the relaxation-to-prior spread method (RTPS; Whitaker and Hamill, 2012) is applied. Covariance localization using a Gaspari-Cohn function (Gaspari and Cohn, 1999) is used for experiments using online-DA and offline-DA. Combination of the relaxation parameter ($\alpha = \{0.0, 0.6, 0.9\}$ in Eq. 2 in Whitaker and Hamill, 2012) and localization scale (radius of influence $ROI = \{2000km, 4000km, 6000km, 8000km\}$) are manually tuned for each experiment so that the analysis root-mean-square error (RMSE) of the SAT is minimized. The experiments with an optimal combination will be shown later.

We use RMSEs to measure the skills of each DA method. We compute RMSEs for the ensemble mean of the time-averaged component ($\langle \bar{\mathbf{x}} \rangle$). In doing so, the area covered by sea-ice are not used to calculate the global mean RMSEs for SST as in such areas SST is output as the average of SAT and SST weighted by sea-ice free and sea-ice covered area, respectively.

Table 1 List of the experiments. Each experiment consists of 7 subexperiments. Each subexperiment assimilates one set of the observations whose averaging time is 1 d, 3 d, 10 d, 1 month, 3 months, 6 months, and 12 months. The variables denoted as T, U, V, Q, Ps, SST, and LST represent temperature, zonal wind, meridional wind, specific humidity, surface pressure, sea surface temperature, and land surface temperature.

Experiment	τ_{OCN} (d)	DA method	Observation	Update variables
NODA	90	-	-	-
OFFLINE	90	Offline-DA	SAT, SST	T, U, V, Q, Ps, SST, LST
ONLINE	90	Online-DA	SAT, SST	T, U, V, Q, Ps, SST, LST
NODA_LONG	360	-	-	-
OFFLINE_LONG	360	Offline-DA	SAT, SST	T, U, V, Q, Ps, SST, LST
ONLINE_LONG	360	Online-DA	SAT, SST	T, U, V, Q, Ps, SST, LST
ONLINE_AA	90	Online-DA	SAT	T, U, V, Q, Ps, LST
ONLINE_OO	90	Online-DA	SST	SST
ONLINE_OA	90	Online-DA	SST	T, U, V, Q, Ps, LST
ONLINE_AO	90	Online-DA	SAT	SST

2.4 Predictability

We measure the predictability of SAT and SST with ensemble forecasts and anomaly correlation coefficient (ACC). The ensemble forecasts are initialized in three different ways where 1) the initial conditions are identical to those for the nature run but with independent small perturbations drawn from the normal distribution with zero mean and standard deviation of 10^{-6}, 2) the initial conditions are drawn from the analysis ensemble members of ONLINE, and 3) the initial conditions are drawn from dates in subsequent years of the nature run, and the time-averaged component of the initial conditions is replaced by that of the nature run and leave the deviation from the time-averaged component unchanged. The first one measures the upper bound of the predictability of SPEEDY. Although a larger ensemble size than that used in this study may result in longer predictability by increasing signal-to-noise ratio (e.g. Scaife et al., 2014), here we refer to the predictability as "intrinsic predictability" (Lorenz, 2006; Melhauser and Zhang, 2012; Zhang et al., 2006) for convenience. The second one measures the predictability of SPEEDY assimilating time-averaged observations with TAU, which will be referred to as “practical predictability” (Lorenz, 2006; Melhauser and Zhang, 2012; Zhang et al., 2006). In this case, errors are in both low- and high-frequency components. The former is from the analysis error, and the latter is because of the fact that TAU does not constrain the high-frequency component (c.f. Section 2.2). The last one measures the predictability when there is no error in the low-frequency component but there is error in the high-frequency component. Thus, it reveals the upper limit of the predictability when TAU is used to initialize the ensemble forecasts in which TAU finds perfect solution for time-averaged component (i.e. same as the truth). Hereafter, we refer to the predictability as “potential predictability”. Intrinsic predictability is “intrinsic” to the system of interest. On the other hand, potential predictability is affected by DA

method (i.e. TAU in this study), and practical predictability is affected by DA experiment settings (e.g. localization, covariance inflation, etc) and the number of and errors in observation as well as DA method. Therefore, we can infer what hampers the practical predictability by comparing these predictabilities. For all the forecasts, the ensemble size is 30, and the forecast length is 5 years. A set of the ensemble forecasts is repeated 60 times with different starting dates of the forecasts. ACC is calculated for the time-averaged component of SAT and SST using the ensemble mean of the forecasts and the nature run. The anomaly is calculated with regard to the climatological seasonality for τ shorter than 12 months. When τ is 12 months, the anomaly is calculated with regard to the climatological mean. All the ensemble forecasts are performed with τ_{OCN} of 90 d and 360 d. In this study, a variable with ACC significantly larger than zero at 95 % confidence level is assumed to be predictable as in the previous climate prediction studies (e.g. Collins, 2002), where a one-sided Student t-test is used for the statistical test.

3 Results and discussion

3.1 Case with τ_{OCN} of 90 d

This section compares online-DA and offline-DA with τ_{OCN} of 90 d. The global mean RMSEs normalized by that of NODA are shown in Figure 2. For SAT, the RMSEs for ONLINE are smaller than those of OFFLINE with $\tau \leq 10$ d ($p < 0.05$). Within this range of τ, the RMSE for ONLINE is smaller than that for OFFLINE by 23 % (1 d) to 3 % (10 d). With $\tau \geq 1$ month, the RMSEs for ONLINE are indistinguishable from those for OFFLINE. For SST, the RMSEs of ONLINE are smaller than those of OFFLINE with $\tau \leq 3$ months ($p < 0.05$) and are indistinguishable from those of OFFLINE with $\tau \geq 6$ months. The result indicates ONLINE outperforms OFFLINE if τ is sufficiently short.

We expected that the skills of ONLINE and OFFLINE decrease as τ lengthens as shown in Pendergrass et al. (2012). Interestingly, however, the skill of OFFLINE increases along with τ for SST. Similarly, the skill of ONLINE for SST does not decrease as τ lengthens with τ of 3 d being remarkably worse skill. These results are owing to relatively small ensemble size to suppress spurious correlations between SST and atmospheric variables including SAT when τ is short. Section 3.4 will investigate the issue in detail.

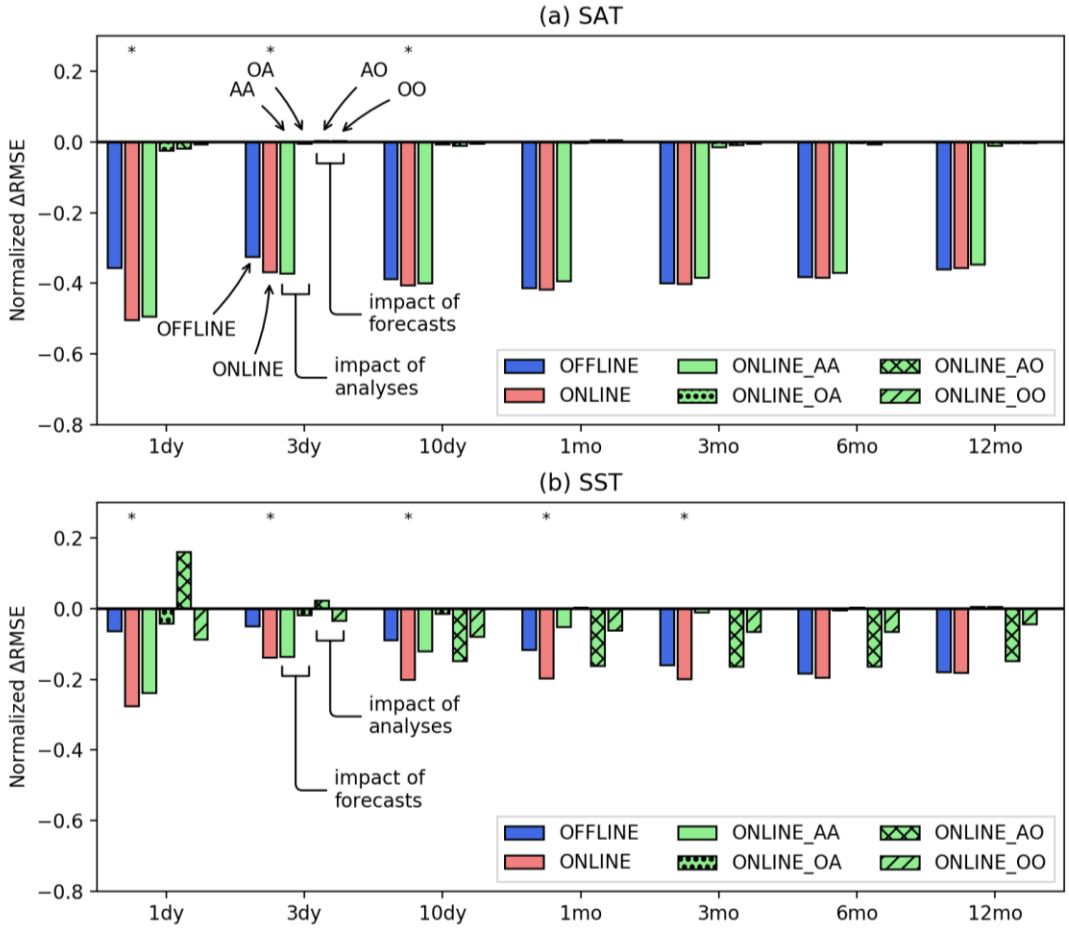

Figure 2 Global mean RMSE difference (XXX-NODA) for (a) SAT and (b) SST, where XXX={OFFLINE, ONLINE, ONLINE_AA, ONLINE_OA, ONLINE_OA, ONLINE_OO}. RMSEs are normalized by the global mean RMSE of NODA for each averaging time τ. Asterisk (*) above red bars denotes RMSE of ONLINE is statistically smaller than that of OFFLINE at significance level of 0.05.

3.2 Predictability

Figure 3 shows global mean ACC for SAT and SST. We determine whether a variable is predictable or not based on ACC. For the intrinsic predictability, the time-averaged SAT and SST are predictable for at least the length of τ for all the τ. For example, 1-month mean SAT is predictable 1 month after the forecasts have started. Here, 1-month lead time means the average of 0-1-month lead time. For the practical predictability, the 1-day, 3-day, and 10-day SAT are predictable during at least 1 d, 3 d, and 10 d after the forecasts are launched, respectively. For SST, experiments other than 12-month mean are predictable for each averaging time after the forecasts have started.

Correspondences can be found among practical predictabilities and the skills of ONLINE relative to OFFLINE (Figures 2 and 3). That is, when the time-averaged state is predictable for the period of averaging time, ONLINE outperforms OFFLINE. For instance, 1-month mean SST is predictable at lead time of 1 month, and the online-DA outperforms offline-DA when assimilating 1-month mean SST. Conversely, the skill of ONLINE is indistinguishable from that of OFFLINE when the time-averaged state is not predictable for the period of τ after the forecast

is launched. The correspondences suggest that the practical predictability controls whether online-DA outperforms offline-DA or not. Although this is what we can anticipate based on the theory, this study is the first to show the relationship between the predictability and the skill of online-DA relative to offline-DA for paleoclimate reconstruction.

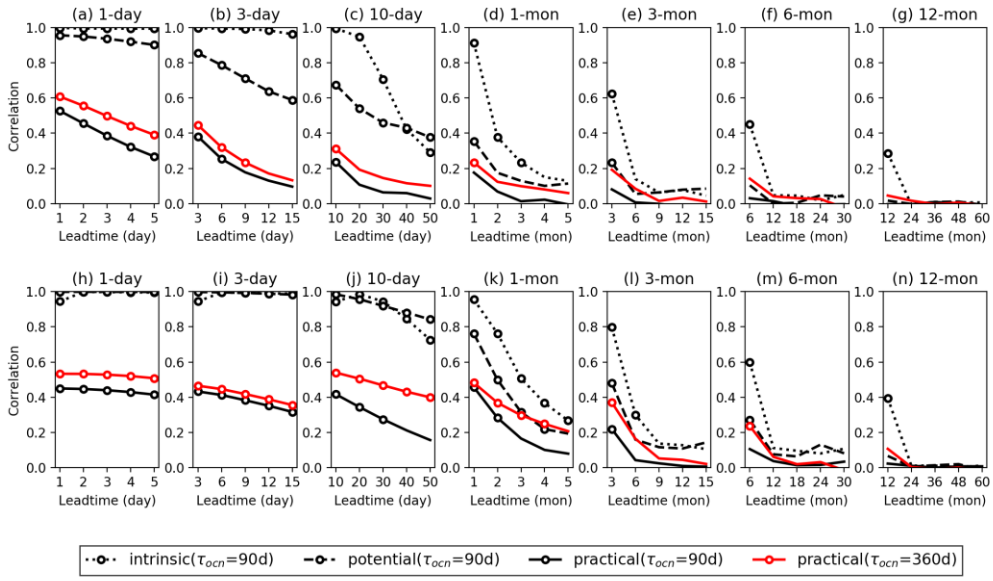

Figure 3 Global mean ACC for (a-g) SAT and (h-n) SST for each averaging time τ. The dotted, dashed, and solid line shows the intrinsic predictability, potential predictability, and practical predictability, respectively. The cases with τ_{ocn} of 90-day are in black and the cases with τ_{ocn} of 360-day are in red. Circles denote that the ACC is statistically higher than zero ($p < 0.05$).

The finding confirms the importance of practical predictability for online-DA to outperform offline-DA. The practical predictability can be further extended by improving DA method, settings, and/or adding more observations and reducing the errors in them. Now, there are large gaps between the intrinsic and practical predictability both for SAT and SST, suggesting large impact the DA method (i.e. TAU), the DA experimental settings (e.g. localization, covariance inflation, etc), and the number of and errors in observation have on the predictability. Here, we briefly investigate which of these limit the practical predictability more strongly using the potential predictability. The potential predictability reveals the upper bound of the predictability with TAU (c.f. Section 2.4). Therefore, the difference between intrinsic and potential predictability shows the extent to which the DA method limits the capability of online-DA, and the difference between the potential and practical predictability shows the extent to which the DA experiment settings and the quantity and quality of the observations limit the capability of online-DA. The comparison of the predictabilities reveals that the loss of practical predictability is mainly due to DA experimental settings when $\tau \leq 10$ d (3 months) and due to DA method when $\tau \geq 1$ month (6 months) for SAT (SST). For instance, for τ of 6 months, the ACC for the intrinsic, potential, and practical predictability is 0.60 ($p < 0.05$), 0.27 ($p < 0.05$), and 0.1, respectively, for SST at the lead-time of 6-month. In this case, almost half of the decrease in ACC is due to the TAU and 2/5 is due to the DA experimental settings and the quantity and quality of the observations. Therefore, although the capability of online-DA is capped by intrinsic nature of the climate system, it is largely limited by the DA method and settings and

observations. In other words, this implies further potency of online-DA if one can come up with better DA method,settings, and/or with more observations and smaller errors in them.

3.3 Case with τ_{OCN} of 360 d

Section 3.2 shows that the online-DA is beneficial when there is long enough practical predictability compared to the averaging time of observations. To further support the claim, we examine how the skill of ONLINE relative to OFFLINE changes with the length of predictability. This experiment would help provide implications on how to translate this study's result into a more realistic case, where the ocean has much longer predictability (e.g. 6-9 years for SAT at annual to subdecadal time scales; Doblas-Reyes et al., 2013) than the slab ocean that we used in this study. Figure 3 shows the predictabilities with τ_{ocn} of 360 d in red. For both intrinsic and practical predictability, the ACCs become stronger, and the predictabilities become longer for both SAT and SST (only practical predictability is shown). The prolonged practical predictability is apparent in 10-day, 1-month, 3-month, and 6-month mean SST.

Figure 4 shows the RMSEs for ONLINE_LONG alongside ONLINE. For SAT, the RMSE remains the same as that of the experiment with τ_{ocn} of 90 d: with $\tau \leq 10$ d, ONLINE_LONG is better than OFFLINE_LONG and they are equally skillful with $\tau \geq 1$ month. On the other hand, for SST, the skills of online-DA relative to offline-DA with τ_{OCN} of 360 d have been improved with τ of 1 month, 3 months, and 6 months, compared to those with τ_{OCN} of 90 d. For τ of 1 month and 3 months, the RMSE of ONLINE_LONG decreased to 0.88 and 0.91, respectively, which were originally 0.92 and 0.96 in ONLINE. For τ of 3 months, the significance level also improved from $p < 0.05$ to $p < 0.001$. Furthermore, ONLINE_LONG outperforms OFFLINE_LONG for τ of 6 months, which was not the case in the experiment with τ_{ocn} of 90 d. Therefore, the result clearly shows the relationship between the predictability and the skill of online-DA relative to offline-DA: the longer the predictability, the more beneficial the online-DA.

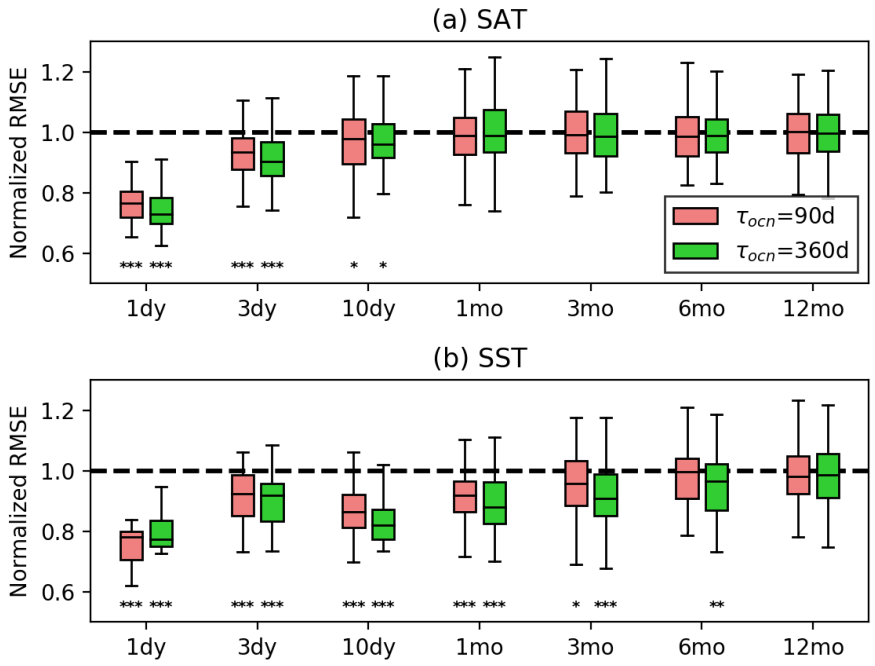

*Figure 4 Global mean RMSE for (a) SAT and (b) SST for each averaging time. Red and green bar is for ONLINE (τ_{ocn}=90 d) and ONLINE_LONG (τ_{ocn}=360 d), respectively. The RMSEs are normalized by global mean of OFFLINE or OFFLINE_LONG. Center lines in the box show the median, the bottom and top of the box shows 25 and 75 percentiles, and whiskers show the range of 1 and 99 percentile of the global mean RMSE. Asterisks *, **, and *** show that the RMSE of ONLINE (ONLINE_LONG) is significantly smaller than that of OFFLINE (OFFLINE_LONG) at the significance level of 0.05, 0.01, and 0.001, respectively.*

3.4 Information source for offline-DA and online-DA

In Section 3.1, we mentioned that the skill of OFFLINE increases along with τ for SST, and the skill of ONLINE with τ of 3 d is notably worse for SST, as opposed to our intuition (Figure 2). The spatial distribution of RMSE would help understand the unexpected results. For OFFLINE with τ of 1 d, the areas where ΔRMSE (OFFLINE-NODA) is negative are confined around SST observations while the ΔRMSE is positive around SAT observations (Figure 5 (a)). Along with τ, the areas where ΔRMSE are negative around SST observations spread, and the areas where ΔRMSE is positive disappear and turn into negative even around SAT observations, suggesting that assimilating SAT is detrimental for SST when τ is short, while it is beneficial when τ is long. Figure 6 shows the correlations between SAT and SST for each τ as a function of their distance. For that, we compute single-point correlations between SAT at each model grid point and SST fields using the nature run. The seasonality is subtracted before the computation. The sample size is 100 for all τ. Figure 6 shows that the correlation between SAT and SST are mostly weak and not statistically significant except for a very close distance (500 km) when $\tau \leq 3$ d. With longer τ, the correlations are stronger. This suggests that the degradation in the skill of OFFLINE with short τ should be due to the weak correlation between SAT and SST and the insufficient ensemble size to suppress the spurious correlations. Indeed, the detrimental impact of DA on SST is largely mitigated in an offline-DA experiment with a 100-member ensemble (not shown).

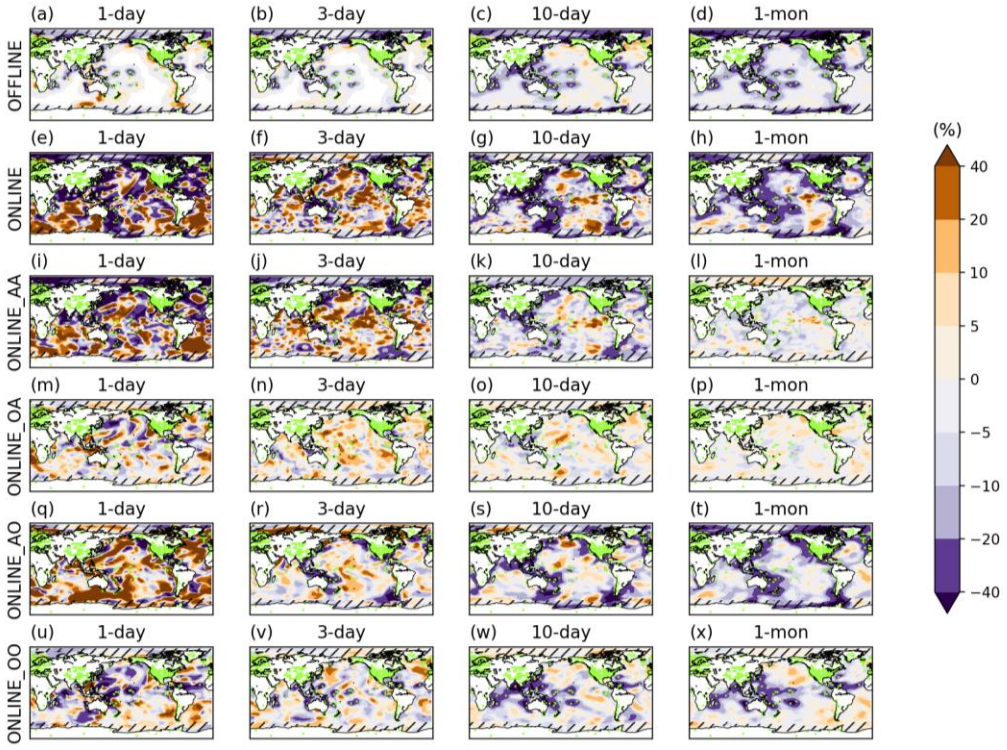

Figure 5 RMSE difference for SST between NoDA and OFFLINE (a-d), ONLINE (e-h), ONLINE_AA(i-l), ONLINE_OA (m-p), ONLINE_AO (q-t), and ONLINE_OO (u-x). Cold (warm) colors show that the experiment is better (worse) than NODA. The RMSE differences are normalized by that of NODA with corresponding averaging time. Green dots denote the observation points. Sea-ice covered area is hatched.

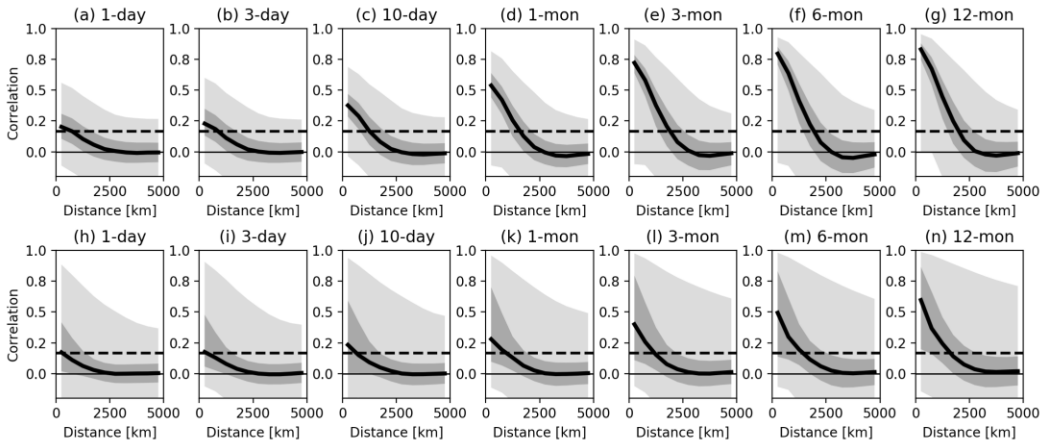

Figure 6 Correlation between SAT and SST for (a-g) SPEEDY and (h-n) a full CGCM named MIROC5 (Watanabe et al., 2010). Solid lines show the median. Dark gray areas show the 25-75 percentile and light gray areas show the 1-99 percentile. Dotted lines denote the correlation coefficient is statistically higher than zero at significance level of 0.05.

For ONLINE, the degraded RMSE compared to NODA is notable when $\tau \leq 3$ d similarly to OFFLINE but with a larger amplitude (Figure 5(e)-(g)). As online-DA transfers either "good" or "bad" information to the subsequent DA-cycles, such information is accumulated in ONLINE, resulting in a larger amplitude in RMSE compared to OFFLINE. The spatial pattern is difficult to readily interpret what degrades the skill of SST in ONLINE as it is affected both in analysis and forecast steps. To disentangle the cause of the degraded skill in SST, we conducted another four experiments with online-DA; ONLINE_AA, ONLINE_OO, ONLINE_AO, and ONLINE_OA (see Sect. 2.3 and Table. 1). Negative impacts of DA are apparent in ONLINE_AA for $\tau \leq 3$ d (Figures 5(i) and (j)). With increasingly larger τ, the negative impact decreases, and it becomes positive overall (Figures 5(k) and (l)). For $\tau \leq 3$ d, the RMSEs are larger than NODA at distance from the SAT observations while they are smaller at around the observations i.e. the skills are degraded over the middle of the oceans and improved along the coast lines. The similar pattern can be found in RMSE for SAT in ONLINE_AA (not shown), which is probably due to the suboptimal localization scale and covariance inflation factor for the area due to the sparse observation, suggesting that the atmosphere disturbs the ocean there. ONLINE_AO similarly exhibits the negative impact of DA for τ of 1 d (Figure 5(q)). This is because of the weak correlation between SAT and SST when τ is small (Figure 6(a)). With increasingly larger τ, the negative impact is softened (Figures 5(r) and (s)), and eventually the impact becomes positive globally (Figure 5(t)) due to the strong correlation between SAT and SST when τ is large (Figure 6(d)). The impact of updating SAT by assimilating SST is marginal for all τ (Figures 5(m)-(p)) as the increments in the atmospheric variables by assimilating SST is small (not shown) probably because the number of SST observations is small. The impact of updating SST by assimilating SST is overall positive, resulting in decreased RMSE in ONLINE_OO globally for all τ (Figures 5(u)-(x)).

The global mean RMSEs for the experiments with online-DA are summarized in Figure 2. For SAT, regardless of τ, most of the improvements in ONLINE comes from SAT observations to update the atmospheric variables, and updating atmospheric and oceanic variables by assimilating SST and updating oceanic variables by assimilating SAT play minor roles. On the

other hand, for SST, each of them plays a considerable role in ONLINE with updating atmospheric variables by assimilating SAT being the largest for $\tau \leq 3$ d, showing that the SST is improved in the forecast steps through heat exchange with the atmosphere. Within the timescales, the impact SAT observations have on SST was negative due to the weak correlation between SAT and SST. With $\tau \geq 1$ month, the benefit of interacting with the atmosphere is small because the atmosphere loses the predictability. Interestingly, within the timescale, the large portion of improvement in SST in ONLINE depends on updating SST with SAT observations. This suggests the importance of the atmospheric observations over land for the ocean in these timescales. It is indicative especially when considering the fact that most of the observations for the last millennium represent SAT in reality (Figure 1 (a)), and the correlations between SAT and SST are fairly large at these timescales (Figure 6).

3.5 Implications to the real case

We showed in Section 3.3, the longer the predictability, the more beneficial the online-DA compared to offline-DA when assimilating observations whose temporal resolution is shorter than the length of predictability. Now, in the real case (i.e. real paleoclimate observations and a more realistic CGCM), temporal resolution of the paleoclimate proxies for the last millennium is most typically annual. On the other hand, annual mean surface temperature is practically predictable for several years (e.g. Collins, 2002; Chikamoto et al., 2013; Doblas-Reyes et al., 2013). Given that the online-DA is beneficial when the predictability is longer than averaging time (i.e. temporal resolution of the observations), the results imply that the online-DA should outperform offline-DA even in the real case.

We also showed that in the experiment with online-DA with $\tau \geq 1$ month, the skill for SST is largely owing to the SAT observations in updating the oceanic variables in analysis steps. In turn, this is due to the strong correlation between SAT and SST and the fact that bulk of the observations represent SAT. Figure 6 shows the correlation between SAT and SST in a more sophisticated CMIP5-class CGCM named MIROC5 (Watanabe et al., 2010). To compute the correlation, simulations from the preindustrial control experiment are used (see Taylor et al., 2012 for detailed information on the experimental design). High correlations between SAT and SST with $\tau \geq 1$ month compared to those for shorter timescales can be also found in MIROC5 (Figure 6(h)-(n)). Therefore, it is legitimate to expect that the ocean state be improved with SAT observations even with a more sophisticated CGCM with sufficiently long τ.

Lastly, it is important to note that online-DA should adopt SCDA to effectively use the atmospheric observations over land. Otherwise, SST will not be updated by SAT observations in the analysis steps, and the information contained in SAT observations will be lost before it affects to the ocean during the model integration because of the loss of predictability in the atmosphere.

4 Summary

This study compares the skill of online-DA and offline-DA when assimilating time-averaged data using an intermediate complexity AGCM, known as SPEEDY, coupled with a slab ocean model under a perfect model OSSE scenario. We found that online-DA outperforms offline-DA

546 if there is practical predictability longer than temporal resolution of the observations. We further
547 confirmed the results by changing the length of the predictability of the ocean in the SPEEDY
548 model. We found that the longer the predictability, the more beneficial the online-DA compared
549 to offline-DA. This study shows for the first time the superiority of online-DA using a CGCM
550 when assimilating time-averaged data whose averaging time is on the time scales of the
551 paleoclimate data.

552 Considering that the typical temporal resolution of the proxies is annual for the last millennium,
553 and that annual mean surface temperature is practically predictable for several years (e.g. Collins,
554 2002; Chikamoto et al., 2013; Doblas-Reyes et al., 2013) in the real case, we can expect that
555 online-DA improves the skill of paleoclimate reconstruction. Although it is the ocean that plays a
556 vital role in carrying the information contained in the observations over time, we found the
557 importance of the atmospheric observations over land in updating the ocean variables in the
558 analysis steps. Therefore, the online-DA should adopt atmosphere-ocean strongly coupled-DA
559 (SCDA) to effectively use the information contained in observations.

560 On the other hand, when assimilating observations whose averaging time is equal to or shorter
561 than 3 d, SCDA resulted in degraded skills for SST. This is due to the limited ensemble size and
562 low signal-to-noise ratio in the background error covariance. The degraded skill when ensemble
563 size is small is commonly seen in SCDA (e.g. Han et al., 2013; Yoshida and Kalnay, 2018). In
564 such a case, in general, localizing the background error covariance is essential to obtain a better
565 analysis. Indeed, the degradation in SST can be mitigated with tighter localization scale in SCDA
566 (not shown), which indicates that tuning of localization scale for cross-domain covariance or
567 correlation cutoff method as suggested by Yoshida and Kalnay (2018) should be effective. This
568 study also confirms that time-averaging strategy is effective for SCDA as shown by Lu et al.
569 (2015a; 2015b).

570 While this study shows the superiority of online-DA to offline-DA for SST, the benefit of online-
571 DA for the atmospheric variables are found to be limited. In reality, the atmosphere can be
572 influenced by predictable slowly changing SST in the Pacific and the Atlantic e.g. ENSO
573 (Trenberth and Caron, 2000) and Atlantic Multidecadal Oscillation (AMO; Knight et al., 2006).
574 One possible reason of the limited impact on the atmosphere is that such internal variabilities are
575 lacked in the current configuration of SPEEDY in which the model is coupled to the slab ocean.
576 It is also unclear how well the variability in subsurface ocean, where prominent predictability
577 resides (e.g. Chikamoto et al., 2013), can be reconstructed using online-DA assimilating only
578 surface variables, which most proxies represent. In the future, we will investigate the impact
579 online-DA have on the atmosphere and the subsurface ocean using a more sophisticated CGCM.

580 This study assumed the perfect model for all the experiments where model error is excluded. In
581 reality, however, model error has been a major issue in atmosphere-ocean coupled-DA and
582 decadal prediction (Meehl et al., 2009; 2014; WMO, 2017 and references therein). To realize
583 online-DA with a CGCM, it is also an important task to develop an effective way to reduce the
584 biases.

585

586 **Acknowledgments, Samples, and Data**

587 The work was supported by JSPS via KAKEN (17K14397, 18H03794),RIKEN via a Grant-in-
588 Aid for Special Postdoctoral Researchers, and Penn State Center for Advanced Data

Assimilation and Predictability Techniques (ADAPT). We thank Dr. Fred Kucharski for kindly sharing the SPEEDY model. CMIP5 outputs are available at https://esgf-node.llnl.gov/projects/cmip5/. All the programs except for those of SPEEDY are available at https://github.com/ats-okazaki/Okazaki-et-al_2020_JGR-A.

References

Abid, M. A., Kang, I. S., Almazroui, M, & Kucharski, F. (2015), Contribution of synoptic transients to the potential predictability of PNA circulation anomalies: El Niño versus La Niña. *Journal of Climate*, 28(21), 8347-8362. doi:10.1175/JCLI-D-14-00497.1

Acevedo, W., Fallah, B., Reich, S., & Cubasch, U. (2017), Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model. *Climate of the Past*, 13, 545–557, doi:10.5194/cp-13-545-2017

Amezcua, J., Ide, K., Kalnay, E., & Reich, S. (2014), Ensemble transform Kalman–Bucy filters. *Quarterly Journal of the Royal Meteorological Society*, 140, 995-1004. doi:10.1002/qj.2186

Annan, J. D., & Hargreaves, J. C. (2012), Identification of climatic state with limited proxy data. *Climate of the Past*, 8, 1141–1151. https://doi.org/10.5194/cp-8-1141-2012

Bahaga, T.K., Mengistu Tsidu, G., Kucharski, F., & Diro, G.T. (2015), Potential predictability of the sea-surface temperature forced equatorial East African short rains interannual variability in the 20th century. *Quarterly Journal of the Royal Meteorological Society*, 141, 16-26. doi:10.1002/qj.2338

Bhend, J., Franke, J., Folini, D., Wild, M., & Brönnimann, S. (2012), An ensemble-based approach to climate reconstructions. *Climate of the Past*, 8, 963–976. doi:10.5194/cp-8-963-2012

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

Collins, M. (2002), Climate predictability on interannual to decadal time scales: the initial value problem. *Climate Dynamics*, 19, 671–692. doi:10.1007/s00382-002-0254-8

Dee, S. G., Steiger, N. J., Emile-Geay, J., & Hakim, G. J. (2016), On the utility of proxy system models for estimating climate states over the common era. *Journal of Advances in Modeling Earth Systems*, 8, 1164– 1179, doi:10.1002/2016MS000677

Dirren, S., & Hakim, G. J. (2005), Toward the assimilation of time-averaged observations. *Geophysical Research Letters*, 32, L04804, doi:10.1029/2004GL021444

Doblas-Reyes, F., Andreu-Burillo, I., Chikamoto, Y., Garcia-Serrano, J., Guemas, V., Kimoto, M. et al. (2013), Initialized near-term regional climate change prediction. *Nature Communications*, 4, 1715. doi:10.1038/ncomms2704

Ehsan, M.A., Kang, I., Almazroui, M., Abid, M. A., & Kucharski, F. (2013), A quantitative assessment of changes in seasonal potential predictability for the twentieth century. *Climate Dynamics*, 41, 2697–2709. doi:10.1007/s00382-013-1874-x

627 Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using
628 Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5),
629 10143-10162. doi:10.1029/94JC00572

630 Franke, J., Brönnimann, S., Bhend, J., & Brugnara, Y. (2017), A monthly global paleo-reanalysis
631 of the atmosphere from 1600 to 2005 for studying past climatic variations. *Scientific Data*, 4,
632 170076. doi:10.1038/sdata.2017.76

633 Gaspari, G., & Cohn, S.E. (1999), Construction of correlation functions in two and three
634 dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723-757.
635 doi:10.1002/qj.49712555417

636 Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., & Mann, M. E. (2006), Using
637 paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the
638 climate of the past millennium. *Climate Dynamics*, 27, 165–184. doi:10.1007/s00382-006-0128-
639 6

640 Goosse, H., Crespin, E., de Montety, A., Mann, M. E., Renssen, H., & Timmermann, A. (2010),
641 Reconstructing surface temperature changes over the past 600 years using climate model
642 simulations with data assimilation. *Journal of Geophysical Research*, 115, D09108.
643 doi:10.1029/2009JD012737

644 Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallz-Damaz,
645 Y., & Shindell, D. (2012a), The role of forcing and internal dynamics in explaining the
646 "Medieval Climate Anomaly". *Climate Dynamics*, 39, 2847–2866. doi:10.1007/s00382-012-
647 1297-0

648 Goosse, H. Guiot, J., Mann, M. E., Dubinkina, S., & Sallaz-Damaz, Y. (2012b), The medieval
649 climate anomaly in Europe: Comparison of the summer and annual mean signals in two
650 reconstructions and in simulations with data assimilation. *Global and Planetary Change*, 84-85,
651 35-47, doi:10.1016/j.gloplacha.2011.07.002

652 Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt. B. R. (2011) Balance and Ensemble
653 Kalman Filter Localization Techniques. *Monthly Weather Review*, 139, 511-
654 522. doi:10.1175/2010MWR3328.1

655 Grottoli, A. G. (2006), Monthly resolved stable oxygen isotope record in a Palauan sclerosponge
656 *Acanthocheatetes wellsi* for the period of 1977-2011. *Proceedings of 10th International Coral*
657 *Reef Symposium*, 572-579.

658 Hakim, G. J. (2000), Role of nonmodal growth and nonlinearity in cyclogenesis initial-value
659 problems. *Journal of the Atmospheric Sciences*, 57, 2951-2967.

660 Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., &
661 Perkins, W. A. (2016), The last millennium climate reanalysis project: Framework and first
662 results. *Journal of Geophysical Research-Atmospheres*, 121, 6745– 6764,
663 doi:10.1002/2016JD024751

664 Han, G., Wu, X., Zhang, S., Liu, Z., & Li, W. (2013), Error Covariance Estimation for Coupled
665 Data Assimilation Using a Lorenz Atmosphere and a Simple Pycnocline Ocean Model. *Journal*
666 *of Climate*, 26, 10218–10231. doi:10.1175/JCLI-D-13-00236.1

667 Hatfield, S. E., Düben, P. D., Chantry, M., Kondo, K., Miyoshi, T., & Palmer, T. N. (2018),
668 Choosing the optimal numerical precision for data assimilation in the presence of model error.
669 *Journal of Advances in Modeling Earth Systems*, 10, 2177– 2191. doi:10.1029/2018MS001341

670 Houtekamer, P. L., & Zhang, F. (2016), Review of the Ensemble Kalman Filter for Atmospheric
671 Data Assimilation. *Monthly Weather Review*, 144, 4489–4532. doi:10.1175/MWR-D-15-0440.1

672 Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007), Efficient data assimilation for
673 spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear*
674 *Phenomena*, 230(1-2), 112-126. doi:10.1016/j.physd.2006.11.008

675 Huntley, H.S., & Hakim, G.J. (2010), Assimilation of time-averaged observations in a quasi-
676 geostrophic atmospheric jet model. *Climate Dynamics*, 35, 995–1009. doi:10.1007/s00382-009-
677 0714-5

678 Kalnay, E. (2003). Atmospheric Modeling, Data Assimilation and Predictability. New York:
679 Cambridge University Press.

680 Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. & Ballabrera-Poy, J. (2007), 4-D-Var or ensemble
681 Kalman filter? *Tellus A: Dynamic Meteorology and Oceanography*, 59(5), 758-773.
682 doi:10.1111/j.1600-0870.2007.00261.x

683 Knight, J. R., Folland, C. K., & Scaife, A. A. (2006), Climate impacts of the Atlantic
684 Multidecadal Oscillation. *Geophysical Research Letters*, 33, L17706.
685 doi:10.1029/2006GL026242

686 Kondo, K., & Miyoshi, T. (2016), Impact of removing covariance localization in an ensemble
687 Kalman filter: experiments with 10,240 members using an intermediate AGCM. *Monthly*
688 *Weather Review*, 144, 4849-4865. doi:10.1175/MWR-D-15-0388.1

689 Kondo, K., & Miyoshi, T. (2019), Non-Gaussian statistics in global atmospheric dynamics: a
690 study with a 10240-member ensemble Kalman filter using an intermediate AGCM. *Nonlinear*
691 *Processes in Geophysics*, 26, 211-225. doi:10.5194/npg-26-211-2019

692 Kotsuki, S., Pensoneault, A., Okazaki, A. and Miyoshi, T. (2020), Weight Structure of the Local
693 Ensemble Transform Kalman Filter: A Case with an Intermediate AGCM. *Quarterly Journal of*
694 *the Royal Meteorological Society*, Accepted Author Manuscript. doi:10.1002/qj.3852

695 Kucharski, F., Molteni, F. & Bracco, A. (2006), Decadal interactions between the western
696 tropical Pacific and the North Atlantic Oscillation. *Climate Dynamics*, 26, 79–91.
697 doi:10.1007/s00382-005-0085-5

698 Kurahashi-Nakamura, T., Paul, A., & Losch, M. (2017), Dynamical reconstruction of the global
699 ocean state during the Last Glacial Maximum. *Paleoceanography*, 32, 326– 350.
700 doi:10.1002/2016PA003001

701 Li, H., Kalnay, E., & Miyoshi, T. (2009), Simultaneous estimation of covariance inflation and
702 observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal*
703 *Meteorological Society*, 135, 523-533. doi:10.1002/qj.371

704 Lorenz, E. N. (1963), Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20,
705 130-141.

706 Lorenz, E. N. (1969), The predictability of a flow which possesses many scales of motion. *Tellus*,
707 21(3), 289-307. doi:10.3402/tellusa.v21i3.10086

708 Lorenz, E. N. (2006), Predictability – a problem partly solved. In T. Palmer and R. Hagedorn
709 (Eds.), Predictability of Weather and Climate (pp. 40-58). New York: Cambridge University
710 Press.

711 Lu, F., Liu, Z., Zhang, S., & Liu, Y. (2015a), Strongly Coupled Data Assimilation Using
712 Leading Averaged Coupled Covariance (LACC). Part I: Simple Model Study. *Monthly Weather*
713 *Review*, 143, 3823–3837. doi:10.1175/MWR-D-14-00322.1

714 Lu, F., Liu, Z., Zhang, S., Liu, Y., & Jacob, R. (2015b), Strongly Coupled Data Assimilation
715 Using Leading Averaged Coupled Covariance (LACC). Part II: CGCM Experiments. *Monthly*
716 *Weather Review*, 143, 4645–4659. doi:10.1175/MWR-D-15-0088.1

717 Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., & Ni, F.
718 (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations
719 over the past two millennia. *Proceedings of the National Academy of Sciences*, 105 (36) 13252-
720 13257, doi:10.1073/pnas.0805721105

721 Mathiot, P., Goosse, H., Crosta, X., Stenni, B., Braida, M., Renssen, H., Van Meerbeeck, C. J.,
722 Masson-Delmotte, V., Mairesse, A., & Dubinkina, S. (2013), Using data assimilation to
723 investigate the causes of Southern Hemisphere high latitude cooling from 10 to 8 ka BP. *Climate*
724 *of the Past*, 9, 887–901. doi:10.5194/cp-9-887-2013

725 Matsikaris, A., Widmann, M., & Jungclaus, J. (2015), On-line and off-line data assimilation in
726 palaeoclimatology: a case study. *Climate of the Past*, 11, 81–93. doi:10.5194/cp-11-81-2015

727 Meehl, G. A., Goddard, L., Murphy, J, Stouffer, R. J., Boer, G., Danabasoglu, G. et al. (2009),
728 Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, 90(10),
729 1467– 1485. doi:10.1175/2009BAMS2778.1

730 Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Cassou, C., Corti, S. et al. (2014), Decadal
731 climate prediction: An update from the trenches. *Bulletin of the American Meteorological*
732 *Society*, 95, 243– 267. doi:10.1175/BAMS-D-12-00241.1

733 Melhauser, C., & Zhang, F. (2012), Practical and Intrinsic Predictability of Severe and
734 Convective Weather at the Mesoscales. *Journal of the Atmospheric Sciences*, 69, 3350–3371.
735 doi:10.1175/JAS-D-11-0315.1

736 Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model.
737 Ph.D. dissertation, University of Maryland, College Park, 197pp.

738 Miyoshi, T. (2011), The Gaussian Approach to Adaptive Covariance Inflation and Its
739 Implementation with the Local Ensemble Transform Kalman Filter. *Monthly Weather*
740 *Review*, 139, 1519-1535. doi:10.1175/2010MWR3570.1

741 Miyoshi, T., Kondo, K., & Imamura, T. (2014), The 10,240-member ensemble Kalman filtering
742 with an intermediate AGCM. *Geophysical Research Letters*, 41, 5264– 5271,
743 doi:10.1002/2014GL060863

744 Molteni, F. (2003), Atmospheric simulations using a GCM with simplified physical
745 parametrizations. I: model climatology and variability in multi-decadal experiments. *Climate*
746 *Dynamics*, 20, 175–191. doi:10.1007/s00382-002-0268-2

747 Okazaki, A., & Yoshimura, K. (2017), Development and evaluation of a system of proxy data
748 assimilation for paleoclimate reconstruction. *Climate of the Past*, 13, 379–393. doi:10.5194/cp-
749 13-379-2017

750 Okazaki, A., & Yoshimura, K. (2019), Global evaluation of proxy system models for stable
751 water isotopes with realistic atmospheric forcing. *Journal of Geophysical Research:*
752 *Atmospheres*, 124, 8972– 8993. doi:10.1029/2018JD029463

753 Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., & Kosro, P. M. (2002), Assimilation of
754 surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical*
755 *Research*, 107(C9), 3122. doi:10.1029/2000JC000511

756 Oke, P.R., Schiller, A., Griffin, D.A. and Brassington, G.B. (2005), Ensemble data assimilation
757 for an eddy-resolving ocean model of the Australian region. *Quarterly Journal of the Royal*
758 *Meteorological Society*, 131, 3301-3311. doi:10.1256/qj.05.95

759 PAGES 2k Consortium (2013), Continental-scale temperature variability during the past two
760 millennia. *Nature Geoscience*, 6, 339–346. doi: 10.1038/ngeo1797

761 Pendergrass, A. G., Hakim, G. J., & Battisti, D. S. (2012), Coupled air-mixed layer temperature
762 predictability for climate reconstruction. *Journal of Climate*, 25, 459-472.
763 doi:10.1175/2011JCLI4094.1

764 Penny, S. G., and T. M. Hamill, 2017: Coupled Data Assimilation for Integrated Earth System
765 Analysis and Prediction. *Bulletin of the American Meteorological Society*, 98, ES169–
766 ES172. doi:10.1175/BAMS-D-17-0036.1.

767 Perkins, W. A. & Hakim, G. J. (2017), Reconstructing paleoclimate fields using online data
768 assimilation with a linear inverse model. *Climate of the Past*, 13, 421–436. doi:10.5194/cp-13-
769 421-2017

770 Renssen, H., Mairesse, A., Goosse, H., Mathiot, P, Heiri, O., Roche, D. M. et al. (2015),
771 Multiple causes of the Younger Dryas cold period. *Nature Geoscience*, 8, 946–949.
772 doi:10.1038/ngeo2557

773 Sluka, T. C., Penny, S. G., Kalnay, E., & Miyoshi, T. (2016), Assimilating atmospheric
774 observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical*
775 *Research Letters*, 43, 752– 759. doi:10.1002/2015GL067238

776 Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., & Roe, G. H. (2014), Assimilation of
777 time-averaged pseudoproxies for climate reconstruction. *Journal of Climate*, 27, 426-441.
778 doi:10.1175/JCLI-D-12-00693.1

779 Steiger, N. J., Smerdon, J. E., Cook, E. R., & Cook, B. I. (2018), A reconstruction of global
780 hydroclimate and dynamical variables over the Common Era. *Scientific Data*, 5, 180086.
781 doi:10.1038/sdata.2018.86

782 Tardif, R., Hakim, G.J. & Snyder, C. (2014), Coupled atmosphere–ocean data assimilation
783 experiments with a low-order climate model. *Climate Dynamics*, 43, 1631–1643.
784 doi:10.1007/s00382-013-1989-0

785 Tardif, R., Hakim, G.J. & Snyder, C. (2015), Coupled atmosphere–ocean data assimilation
786 experiments with a low-order model and CMIP5 model data. *Climate Dynamics*, 45, 1415–1427.
787 doi:10.1007/s00382-014-2390-3

788 Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D.
789 M., Steig, E. J., & Noone, D. (2019), Last Millennium Reanalysis with an expanded proxy
790 database and seasonal proxy modeling. *Climate of the Past*, 15, 1251–1273, doi:10.5194/cp-15-
791 1251-2019

792 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012), An Overview of CMIP5 and the
793 experiment design. *Bulletin of the American Meteorological Society*, 93, 485-498.
794 doi:10.1175/BAMS-D-11-00094.1

795 Trenberth, K. E., & Caron, J. M. (2000), The southern oscillation revisited: Sea level pressures,
796 surface temperatures and precipitation. *Journal of Climate*, 13, 4358–4365, doi:10.1175/1520-
797 0442(2000)013<4358:TSORSL>2.0.CO;2

798 van der Schrier, G., & Barkmeijer, J. (2005), Bjerknes' hypothesis on the coldness during AD
799 1790–1820 revisited. *Climate Dynamics*, 25, 537–553. doi:10.1007/s00382-005-0053-0

800 van Leeuwen, P. J. (2009), Particle filtering in geophysical systems. *Monthly Weather Review*,
801 137, 4089-4114. doi:10.1175/2009MWR2835.1

802 von Storch, H., Cusbach, U., Gonzalez-Rouco, J. F., Jones, J. M., Voss, R., Widmann, M., &
803 Zorita, E. (2000), Combining paleoclimatic evidence and GCMs by means of data assimilation
804 through upscaling and nudging (DATUN). *11th American Symposium on Global Chance*, 28-31.
805 Available at http://www.hvonstorch.de/klima/pdf/DATUN-2000.pdf

806 Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T.,
807 Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohota, T., Nozawa, T.,
808 Hasumi, H., Tatebe, H., & Kimoto, M. (2010), Improved climate simulation by MIROC5: Mean
809 States, Variability, and Climate Sensitivity. *Journal of Climate*, 23, 6312–6335.
810 doi:10.1175/2010JCLI3679.1

811 Whitaker, J. S. & Hamill, T. M. (2012), Evaluating methods to account for system errors in
812 ensemble data assimilation. *Monthly Weather Review*, 140, 3078-3089. doi:10.1175/MWR-D-
813 11-00276.1

814 WMO (2017), Coupled Data Assimilation for Integrated Earth System Analysis and Prediction:
815 Goals, Challenges and Recommendations. WWRP 2017-3. Available at
816 https://www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf

817 Yoshida, T. & Kalnay, E. (2018), Correlation-Cutoff Method for Covariance Localization in
818 Strongly Coupled Data Assimilation. *Monthly Weather Review*, 146, 2881–
819 2889. doi:10.1175/MWR-D-17-0365.1.

820 Zhang, F., Odins, A. M., & Nielsen-Gammon, J. W. (2006). Mesoscale Predictability of an
821 Extreme Warm-Season Precipitation Event. *Weather Forecasting*, 21, 149-
822 166. doi:10.1175/WAF909.1

823