Benefits of CMIP5 Multimodel Ensemble in Reconstructing Historical Ocean Subsurface Temperature Variations

LIJING CHENG AND JIANG ZHU

International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ABSTRACT

A complete map of the ocean subsurface temperature is essential for monitoring aspects of climate change such as the ocean heat content (OHC) and sea level changes and for understanding the dynamics of the ocean/climate variation. However, global observations have not been available in the past, so a mapping strategy is required to fill the data gaps. In this study, an advanced mapping method is proposed to reconstruct the historical ocean subsurface (0–700 m) temperature field from 1940 to 2014 by using ensemble optimal interpolation with a dynamic ensemble (EnOI-DE) approach and a multimodel ensemble of phase 5 of the Coupled Model Intercomparison Project (CMIP5) historical and representative concentration pathway 4.5 simulations. The reconstructed field is a combination of two parts: a first guess provided by the ensemble mean of CMIP5 models and an adjustment by minimizing the analysis error with the assistance of error covariance determined by the CMIP5 models. The uncertainty of the field can also be assessed. This new approach was evaluated using a series of tests, including subsample tests by using data from the Argo period, idealized tests by specifying a truth field from the models, and withdrawn-data tests by removing 20% of the observations for validation. In addition, the authors showed that the ocean mean state, long-term trends, and interannual and decadal variability are all well represented. Furthermore, the most significant benefit of this method is to provide an improved estimate of the long-term historical OHC changes since 1940, which have important implications for Earth's energy budget.

1. Introduction

The ocean is the key component in tracking Earth's heat budget (Abraham et al. 2013; Allan et al. 2014; Church and White 2011; Church et al. 2011; Trenberth et al. 2014a; Trenberth and Fasullo 2012); therefore, the change in the ocean heat content (OHC) is a critical metric for monitoring ongoing global warming (von Schuckmann et al. 2016). Estimates of historical OHC change rely on in situ temperature profile observations. However, few temperature observations are available spatially or temporally, leaving numerous data gaps. For example, data were available only near the west coast of the Pacific and Atlantic Oceans in 1941 (Fig. 1). The data coverage was extended to the midlatitudes of the northern

Pacific and Atlantic Oceans by 1971, and then by 1991 to most regions of the Northern Hemisphere. Until 2005, observations were scattered globally all over the ocean when the Argo system was inaugurated (with the exception of the Southern Ocean and coastal area) (Freeland et al. 2010; Roemmich et al. 2012; von Schuckmann et al. 2014; Roemmich et al. 2015; Riser et al. 2016). The lack of in situ observations has impeded the accurate assessment of the rate of global warming (Lyman and Johnson 2008; Abraham et al. 2013; Cheng and Zhu 2014a) and the development of an in-depth understanding of ocean and climate dynamics (Gille 2008).

A temperature field with complete spatial and temporal coverage throughout the subsurface is always preferred in oceanographic studies. A large group of methods makes use of data assimilation techniques by taking into account the ocean dynamics represented by ocean models. These efforts have resulted in a number of reanalysis datasets (Balmaseda et al. 2013; Carton and Giese 2008; Chang et al. 2013; Xue et al. 2011). However, OHC estimates in various reanalysis datasets have revealed substantial differences, as indicated by

---

ᵃ Denotes Open Access content.

Corresponding author address: Jiang Zhu, ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, No. 40, Huayanli, Qijiahuozi, Chaoyang, Beijing 100029, China.
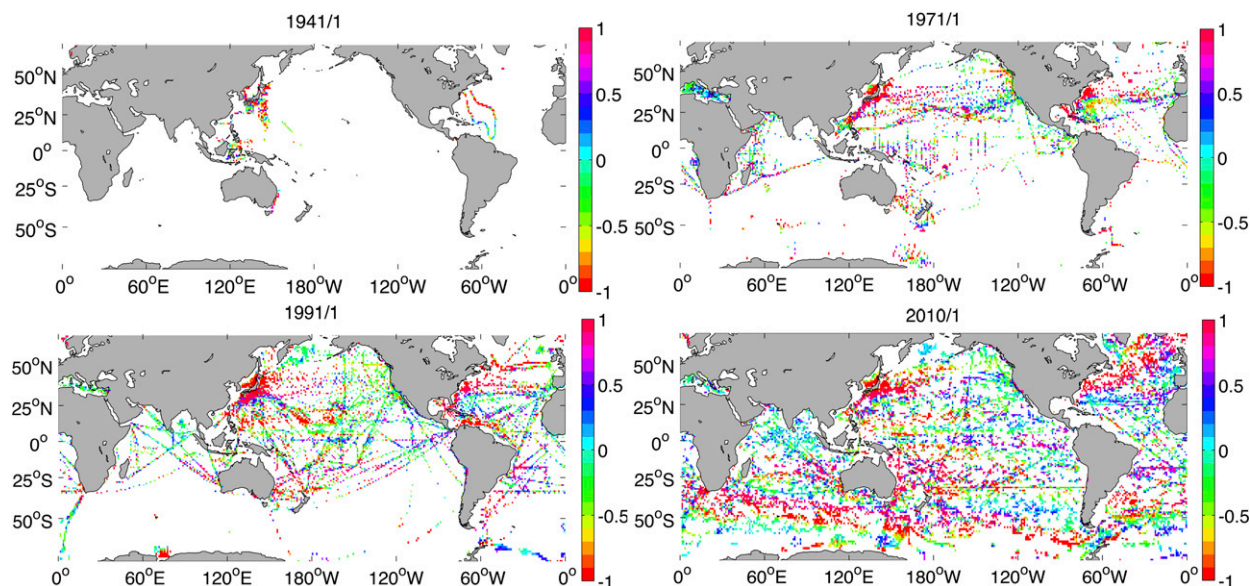E-mail: jzhu@mail.iap.ac.cn

FIG. 1. Coverage of in situ temperature observations at a depth of 10 m in January 1941, 1971, 1991, and 2010. The color shows the average temperature anomaly (°C) in each 1° × 1° grid.

reanalyzed intercomparison studies (Carton and Santorelli 2008; Xue et al. 2012; Palmer et al. 2015).

Another group of strategies has been adopted by the community to statistically fill in the data gaps (Domingues et al. 2008; Good et al. 2013; Ishii et al. 2003; Levitus et al. 2012; Palmer and Haines 2009; Roemmich and Gilson 2009; Smith and Murphy 2007; Willis et al. 2004). Among these methods, Lyman et al. (2010) and Palmer and Haines (2009) used the average temperature anomaly of the sampled areas to represent the global mean, which assumes that ocean change in data-rich regions is the same as that in regions with data gaps. Domingues et al. (2008) adapted an EOF analysis based on the available data and used the principle components obtained to represent the relationship between a data-rich and a data-sparse region. Willis et al. (2004) projected the sea level change onto the OHC field based on their correlations.

In addition to the methods described above, other methods have been used, which are often termed "objective interpolation methods." These techniques require an error covariance field to propagate information from observed regions to regions with data gaps combined with a first-guess field to provide prior information. The first guess is always based on climatology (i.e., with zero anomaly) for simplification (i.e., Levitus et al. 2012; Smith and Murphy 2007; Ishii et al. 2003; Good et al. 2013). This choice might cause the final field to revert to the first-guess field, especially in areas where there are large data gaps, such as in the Southern Hemisphere (in Fig. 1 at 1941, 1971, and 1991). This

problem is suspected to result in underestimated long-term OHC trend (Cheng and Zhu 2014a; Durack et al. 2014a). Therefore, a more appropriate strategy is to improve the first guess; such an improvement is the motivation for this study.

Meanwhile, the error covariance defines the correlations among different ocean grids and determines how the information propagates from data-rich to data-sparse regions. The existing methods use parameterized covariance based on some basic assumptions, for example, typical correlation length scales (Good et al. 2013; Ishii et al. 2003; Levitus et al. 2012). The error covariance fields in these studies are empirically constructed and are mainly determined by the spatial distance between the location analyzed and the location observed. A better choice is to use a covariance defined by the output of a high-resolution climate model simulation, which allows a high-resolution representation of error covariance (Smith and Murphy 2007). However, this proposed method relies on the accuracy of the applied ocean model, which always contains a systematic bias to some extent; therefore, a potential improvement is the use of an ensemble of models, which is another motivation of this study.

Here, an ensemble of model outputs from historical runs and representative concentration pathway 4.5 (RCP4.5) of phase 5 of the Coupled Model Intercomparison Project (CMIP5) is used to provide a more reasonable choice for the first-guess field and error covariance statistics. The conclusion of the Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate

Change (IPCC) states that "there is high confidence that many CMIP5 models reproduce the observed increase in ocean heat content since 1960," and many other studies agree (AchutaRao et al. 2007; Gleckler et al. 2012; Flato et al. 2013; Cheng et al. 2015b; Smith et al. 2015). Therefore, an ensemble of models could be a reasonable choice to generate the first-guess field, which is better than using climatology that has a zero temperature anomaly over time. Moreover, by using the ensemble strategy, the impact of model bias can be minimized and a more accurate value for the covariance could potentially be achieved.

This manuscript is structured as follows: the data used for this study is introduced in section 2. The methods and the related choices of parameterizations are introduced in section 3. To determine one of the key parameters and perform a first validation of the method, the data for a recent decade are used in section 4 because the Argo network helps to maintain a dataset with a near-global coverage. The idealized experiments based on the models are outlined in section 5, which illustrates the improvement due to the proposed strategy when using either a perfect or a slightly biased first-guess field. However, the error covariance is exactly known in these idealized tests. In section 6, we further examine the method by using in situ observations for which the error covariance is unknown. We withdrew a portion of the in situ observations as an independent dataset for validation. In section 7, the obtained reconstructed temperature field is compared with other datasets. A brief description of the analysis field is presented in section 8, in which we highlight the historical upper ocean heat content estimate from 1940 to 2014 based on the analyzed field. A discussion of possible future applications is provided in section 9.

2. Data

In situ ocean temperature observations are sourced from the *World Ocean Database 2013* (*WOD13*) from 1940 to 2014, with more than 10 000 000 temperature profiles (Boyer et al. 2013). The quality flags provided by *WOD13* are used to remove erroneous profiles and measurements. An additional standard deviation check is also utilized to remove spurious temperature measurements, whereby measurements more than five standard deviations away from the mean are removed. XBT bias is corrected by using the process of Cheng et al. (2014), and mechanical bathythermograph (MBT) bias is corrected by using the process of Ishii and Kimoto (2009).

Twelve monthly climatologies were constructed by averaging all of the temperature profiles from 1999 to 2005 into 1° × 1° ocean grid boxes and 27 standard depths, from the sea surface to 700 m. The standard depths were as follows: 1 m, 5 m, 10–100 m in 10-m intervals, 120–200 m in 20-m intervals, and 250–700 m in 50-m intervals. Then, a nine-point median filter (around each specific horizontal grid) was applied to the resulting blended fields. Only seven years (within 1999–2005) of data were used for the following reasons:

1) Cheng and Zhu (2015) indicated that a long-term climatology will bias the long-term calculation of OHC, so a climatology constructed using data from a short period was preferred for this study.
2) The CMIP5 models of historical simulation ended in 2005, and then the RCP4.5 simulation began in 2006. However, the successive RCP simulations for nine models were not available.
3) The global OHC time series obtained using the 1999–2005 climatology is very similar to that obtained by using the 2008–12 climatology when using the method of Cheng et al. (2015b).

The climatology is subtracted from each temperature profile to obtain a temperature anomaly profile. Then, the anomaly profiles are averaged in 1° × 1° grid boxes for each month and 27 standard depths to obtain the grid-averaged and monthly temperature anomaly field. Given the slowness of ocean temporal variation, in each month, we use data collected over a 3-month window to calculate the monthly mean, which helps to increase the data coverage without losing the ENSO variation. In this study, the proposed mapping method was applied to this grid-averaged, monthly temperature anomaly field.

The CMIP5 model output of historical runs from 1940 to 2005 and an RCP4.5 scenario from 2006 to 2014 was assembled for this study (Taylor et al. 2012). The historical runs included 40 separate runs and the RCP4.5 projections included 31 separate runs, which are listed in Table 1. We note here that a so-called "climate drift" exists for CMIP5 climate models (Sen Gupta et al. 2013), which is a spurious long-term variation that cannot be explained by internal and external forced variability. According to Durack et al. (2014a), the upper 0–700 m of the OHC is only weakly affected by climate drift; therefore, we did not apply a correction for climate drift in this study.

The data preparation for each CMIP5 model is similar to the observations. In brief, for each model, the temperature field was first interpolated to a 1° × 1° grid and 27 standard depths, and then 12 monthly climatologies were constructed using the 1999–2005 data. The temperature anomaly field was then obtained by subtracting the climatology from the interpolated field. This process was repeated separately for 40 (31 after 2006) times for

TABLE 1. List of CMIP5 models used in this study. There are 17 different models and 40 (31) different ensemble runs in total from 1940 to 2005 (from 2006 to 2014). [CESM1(BGC) is the Community Earth System Model, version 1, with biogeochemistry; additional acronyms are available online at http://www.ametsoc.org/PubsAcronymList.]

	Model name	Modeling center (or group)	Ensemble run	Projection (RCP4.5)
1	HadCM3	Met Office Hadley Centre	R1, R2, R3	R1, R2, R3
2	CESM1(BGC)	Community Earth System Model contributors	R1	R1
3	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	R1, R2, R3	R1
4	BCC_CSM1.1(m)	Beijing Climate Center, China Meteorological Administration	R1, R2, R3	R1
5	GFDL CM2.1	NOAA/Geophysical Fluid Dynamics Laboratory	R1, R2, R3	R1, R2, R3
6	GISS-E2-H	NASA Goddard Institute for Space Studies	R2, R3	R2, R3
7	GISS-E2-R	NASA Goddard Institute for Space Studies	R2, R3	R2, R3
8	NorESM1-M	Norwegian Climate Centre	R2, R3	—
9	CCSM4	National Center for Atmospheric Research	R1, R2, R3	R1, R2, R3
10	MPI-ESM-P	Max Planck Institute for Meteorology	R1	—
11	MPI-ESM-MR	Max Planck Institute for Meteorology	R1, R2, R3	R1, R2, R3
12	MPI-ESM-LR	Max Planck Institute for Meteorology	R1, R2, R3	R1, R2, R3
13	IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace	R1, R2, R3	R1, R2, R3
14	IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace	R1, R2, R3	R1
15	IPSL-CM5B-LR	L'Institut Pierre-Simon Laplace	R3	R1
16	CanESM2	Canadian Centre for Climate Modelling and Analysis	R1, R2, R3	R1, R2, R3
17	BCC_CSM1.1	Beijing Climate Center, China Meteorological Administration	R1	R1

40 (31) models, and then we obtained 40 (31) grid-averaged, monthly temperature anomaly fields, which served as the input for our proposed mapping method.

The use of the temperature anomaly field rather than temperature field in the mapping is emphasized because the spatial and seasonal variation of the ocean (on the order of 1°C from the equator to the polar region and 0.1°C from summer to winter) is much larger than its interannual (on the order of 0.01°C yr^{-1} for global average signals) and multidecadal variation (on the order of 0.001°C yr^{-1} for global average signals). The climate models have substantial differences in simulating the mean ocean state, including spatial and seasonal variation, but they are not our primary focus. Instead, the anomaly field is used to remove the impact of the ocean spatial and seasonal variation.

3. Methods

A framework combining in situ observations and CMIP5 models is required to obtain the final analysis field. In this section, we briefly describe the method used in this study, the ensemble optimal interpolation method with a dynamic ensemble (EnOI-DE), which is based on the Kalman filter (KF; Kalman 1960; Evensen 1994, 2003, 2004). The analysis field $\mathbf{X}^a$ is a linear combination of a prior guess field $\mathbf{X}^b$ (or background field), and in situ observations (as denoted in matrix $\mathbf{y}$):

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{X}^b), \qquad (1)$$

where $\mathbf{H}$ is the transfer matrix from the analysis space to observation space, and the Kalman gain $\mathbf{K}$ is obtained by

the maximum likelihood method by minimizing the analysis error. The Kalman gain is calculated by

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}. \qquad (2)$$

The superscript T denotes the transposition operation; $\mathbf{R}$ is the error covariance of the observations, which is always predefined (a description of the method will be provided later); and $\mathbf{P}^b$ is the error covariance of the background field. This covariance is essential for propagating the signals from the data-rich to the data-sparse regions. Ideally, $\mathbf{P}^b$ has to be calculated explicitly given the model propagation matrix. However, because ocean dynamics are greatly nonlinear, it is impossible to explicitly calculate $\mathbf{P}^b$. Some methods use an empirically parameterized $\mathbf{P}^b$ as an approximation (Good et al. 2013; Ishii et al. 2003; Levitus et al. 2012), but the accuracy of such a method is unclear. Improved strategies include the ensemble Kalman filter (EnKF) method (Evensen 1994, 2004); the ensemble optimal interpolation (EnOI; Evensen 2003); and the ensemble square root filter (EnSRF; Bishop et al. 2001; Sakov and Oke 2008a,b), which use an ensemble of models to represent the background error covariance ($\mathbf{P}^b$). In this way, N ensemble members of the model outputs are recorded in matrix $\mathbf{A}$, which is defined as $\mathbf{A} = (\mathbf{X}_1^b, \mathbf{X}_2^b, \ldots, \mathbf{X}_N^b)$. The ensemble mean of the models is

$$\mathbf{X}^b = \frac{\mathbf{A1}}{N}, \qquad (3)$$

where $\mathbf{1}$ is the unity vector. The model covariance (or background covariance) $\mathbf{P}^b$ is now calculated as

$$\mathbf{P}^b = \frac{\sum_{i=1}^{N} (\mathbf{X}_i^b - \mathbf{X}^b)(\mathbf{X}_i^b - \mathbf{X}^b)^T}{(N-1)}. \tag{4}$$

For the traditional EnKF or EnSRF method, the error covariance is propagated and updated according to the following equation: $\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b$. Thus, the analyses of the next time step use the information from the current time step. However, in this study, the CMIP5 models were not constrained by observations, so they do not carry the covariance information from the previous cycles forward in time. EnKF-like systems that use static or seasonal ensembles are often classified as the EnOI method. Therefore, the method used in this study is characterized as EnOI-DE, and the dynamic ensemble is provided by the unconstrained simulations of the CMIP5 models. The implementation of the method follows Sakov and Oke (2008a,b).

The general framework of the mapping method has now been developed. The reconstructed field ($\mathbf{X}^a$) is the monthly temperature anomaly in each $1° \times 1°$ ocean grid at each standard depth. In other words, the mapping strategy has been applied to the grid-averaged temperature anomalies in a plane with two-dimensional latitude–longitude. The background field ($\mathbf{X}^b$) is the ensemble mean of the 40 pre-2005 CMIP5 models ($N = 40$) and the 31 models from 2006 to 2014 ($N = 31$), and each observed anomaly ($\mathbf{y}$) is the gridded average and monthly in situ observation.

It is still necessary to define some key parameters for the application of this method. For clarity, the parameters are defined as follows.

- The observation error variance $\mathbf{R}$ represents the error of the observations that must be defined prior to the analysis. It consists of both the instrumental error $\mathbf{Re}$ due to instrument inaccuracy and the representativeness error $\mathbf{Rr}$ due to the need to represent the spatial (at $1° \times 1°$ and 1-m standard grid depths) and temporal (1 month) averages from a limited numbered of observations. Both the instrumental error and the representativeness errors are assumed to be uncorrelated between any two different grids; therefore, both $\mathbf{Re}$ and $\mathbf{Rr}$ are diagonal matrices. This assumption may not be correct if the temperature observations in two different grids are made by using the same instruments or on the same cruise, which might lead to a correlation of the errors in the two grids. For instance, most of the XBT data are obtained by analog systems before 1985, which has a larger bias, and by digital systems after 1990 with smaller bias. However, we decided to ignore this correlation because it is difficult to quantify.

According to the discussion, the observational error variance in each spatial ($1° \times 1°$ and standard depth) and temporal (1 month) grid is calculated as follows:

$$\mathbf{R} = \mathbf{Re} + \mathbf{Rr} = \frac{\sum_{j=1}^{M} \mathbf{Ei}_j}{M} + \frac{\sigma^2}{M}, \tag{5}$$

where M observations exist for a given grid cell; $\mathbf{Re}$ in each grid cell is set to the mean of the typical precision of the different instruments used to obtain the data in the cell; and $\mathbf{Ei}$ is the precision of the instrument for each individual observation, which is the variance of the instrumental error and is set to $0.1°C^2$ for XBT, $0.001°C^2$ for CTD, $0.002°C^2$ for Argo, $0.3°C^2$ for MBT, and $0.01°C^2$ for the other instruments according to Abraham et al. (2013). Because $\mathbf{Rr}$ defines the accuracy of the monthly mean value in each cell, it is approximated by the variance of the monthly mean values of the temperature anomalies. The symbol σ^2 represents the variance of the various temperature measurements against the monthly mean value. The data from 2005 to 2013 are used to calculate the value of σ^2 in each grid because of the approximately full global coverage of the data, which are shown in Fig. 2 for depths of 1, 300, and 700 m. The representativeness error $\mathbf{Rr}$ is expected to show flow dependency, that is, the error is expected to be higher in areas where a gradient in the flow rate (speed) is present. The error is larger in regions of higher variability, so more observations are required to represent the mean value. Figure 2 shows a larger representativeness error in the boundary-current regions and near the Antarctic Circumpolar Current (ACC) regions.

- Next, the need for a localization strategy should be determined. This strategy assumes that each observation could only impact a limited region around its location. The application of a localization strategy helps to minimize the impact of an imperfect error covariance, which could propagate the information incorrectly over a much longer distance. In general, an error in the model covariance will initially increase with distance from the analysis point, but it is almost impossible to specify an exact box (its size is determined by the influencing radius) within which the observations could benefit the analysis and outside of which they could not because the truth correlation length is unknown. Previous studies either specified a constant length scale over the
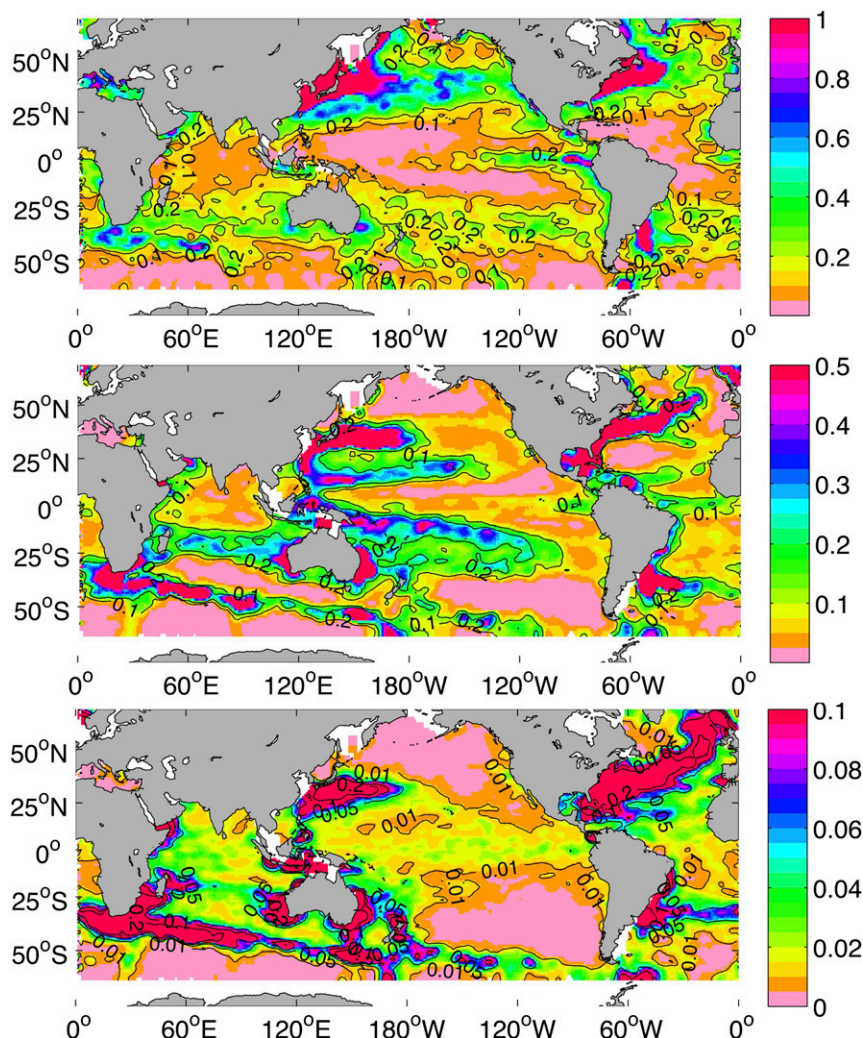
FIG. 2. Representativeness errors in each grid at (top) 10 m, (middle) 300 m, and (bottom) 700 m. The color shows the values of σ^2 in Eq. (5), which were calculated by the standard deviation of different individual temperature measurements against the monthly mean value.

global ocean [i.e., the three-pass analyses of 880, 660, and 440 km in Levitus et al. (2012) or ∼300 and ∼900 km in Willis et al. (2004)] according to the ocean large-scale wavenumber spectral in Zang and Wunsch (2001) or an empirically parameterized scale [i.e., a length scale that decreases with latitude, as in Ishii et al. (2003), or estimates according to the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) data in Smith and Murphy (2007)]. However, a localization strategy interferes with the potential advantage of an improved covariance. Therefore, we included a test in the next section to justify the usage of a localization strategy and empirically determine the influencing radius that defines a specific area; data outside that area are not used for analysis.

4. Evaluation by using data in recent decade

During the last decade, the Argo network achieved near-global coverage (Abraham et al. 2013; Roemmich et al. 2015; Riser et al. 2016). Therefore, in this section, we describe the conduction of a subsample test by sampling the data from the past eight years (2007–14) at historical observation locations. Similar tests were also done by a recent study (Smith et al. 2015). This test was used first to examine the impact of a localization strategy, to determine the maximum influencing radius as discussed in the previous section, and then to preliminarily evaluate the proposed mapping method.

In the subsample test, we used the temperature anomaly at 10 m for each selected month (January and August from 2007 to 2014) as the truth value (for a total
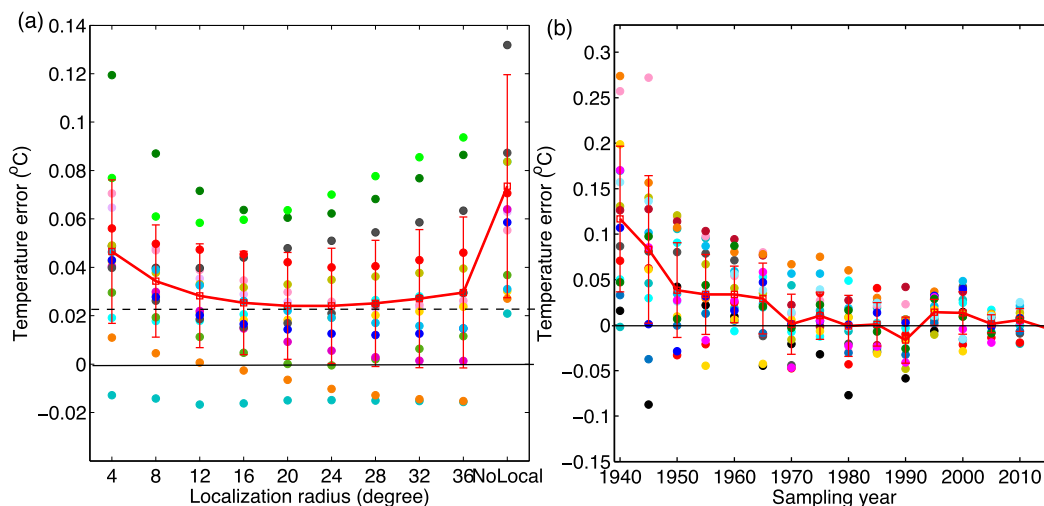
FIG. 3. (a) Mean temperature error as a function of different choices of the influencing radius between the reconstructed and truth fields. Each dot represents the averaged temperature error for each truth field and each influencing radius, where the errors at 16 sampling years are averaged together. (b) Mean temperature error as a function of sampling years from 1941 to 2014, with an influencing radius of 20°. Dots of different colors represent the 16 different truth fields. The red line with squares and the error bars show the mean and standard deviation, respectively.

of 16 truth fields). To achieve better spatial coverage, the data within 3 months around the selected month were averaged together. Each truth field was then subsampled according to the history of the observation locations (on January at every five years from 1941 to 2014). Thus, 16 subsampled fields were generated for each truth field. To examine the performance of the new method, all of the subsampled fields were mapped using the proposed method and then compared with the truth fields. This test was run multiple times by using different influencing radii from 4° to 36° with 4° increments, and another run without using a localization strategy was also carried out.

The global mean error between the mapped fields and the truth fields are shown by dots in Fig. 3a as a function of the influencing radius. It appears that a radius between 16° and 24° minimizes the error of the reconstructed field. An influencing radius larger than 24° allows more remote correlations, but it increases the error, suggesting the inaccuracy of the remote covariance represented by models. An influencing radius less than 16° also increases the analysis error because of the ineffective propagation of information from data-rich to data-sparse regions. Therefore, we used 20° as the influencing radius for the localization strategy. It is noted here that it requires a taper function in the localization strategy, which defines how much an observation can impact the analyzed grid. In this study, we adapted a Gaussian function assuming that the impact of the observation is exponentially decreasing with distance

away from the analyzed grid. Tests to find the best selection of the taper function will be the first priority for the future improvement of the proposed method, for example, the functions proposed by Gaspari and Cohn (1999).

Figure 3a shows that the global mean temperature errors (defined as analysis minus observation) are always positive, indicating an error in the reconstructed field. To further investigate this error, we presented the temperature error as a function of the sampling year from 1941 to 2014 in Fig. 3b, with the influencing radius set to 20°. A positive error was found for the pre-1965 ocean sampling. By contrast, for post-1965 ocean samplings, the proposed mapping method could reconstruct the truth field in an unbiased manner, leading to errors in values around zero. This indicated that the new method could reduce the impact of the change of the observation system from the traditional ship-based system to the Argo system at the beginning of this century, as documented in Cheng and Zhu (2014a). Next, we will explore the reasons that some pre-1965 errors might be present and how well the information was propagated from the data-rich to data-sparse areas.

The geographical distribution of temperature anomalies at 10 m in January 2014 is shown in Fig. 4, along with the reconstructed field. The subsampled fields for each observation location in January 1941, 1961, 1981, and 2001 as well as the corresponding mapped fields are also shown. In 2014, the ocean evidenced a moderate La Niña pattern, with cooling signals in the eastern and
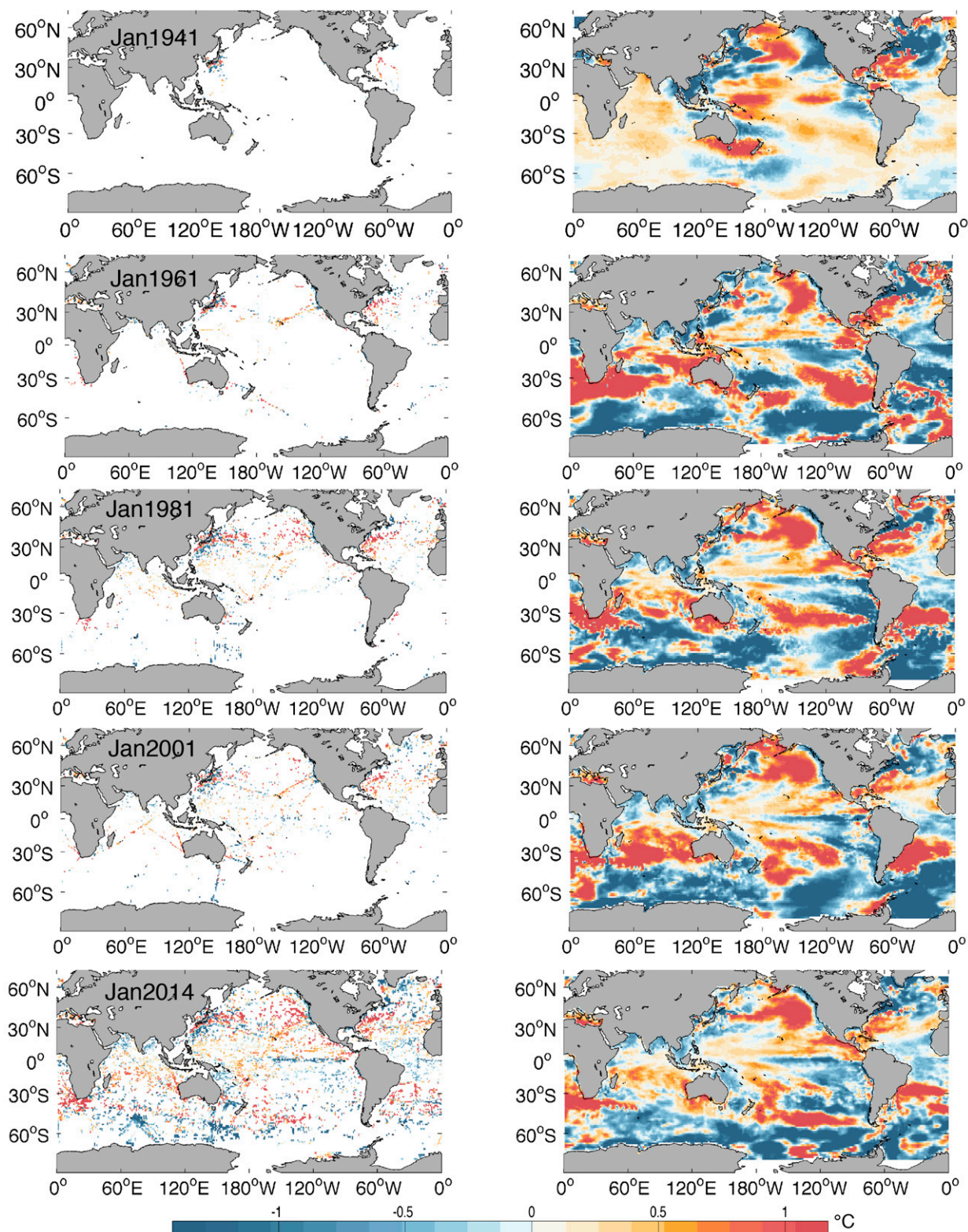
FIG. 4. (left) Subsampled temperature anomalies. The temperature anomaly field at 10 m in January 2014 was subsampled according to the observation locations (top)–(bottom) in January 1941, 1961, 1981, 2001, and 2014. The color shows the average temperature anomaly in each $1° \times 1°$ grid. (right) The field mapped by using the proposed method is presented.

southeastern Pacific together with strong warm anomalies in the northwestern Pacific. The Indian Ocean experienced cooling in the north and warming in the south, and the Atlantic Ocean showed warming in the midlatitudes and cooling in the other latitudes.

In January 1941, only some grids were observed in the coastal regions near Japan, western Australia, and in the northwestern Atlantic Ocean, so the temperature anomaly pattern in midlatitudes in the Northern Hemisphere and near Australia can be correctly reconstructed. However, in the western Indian, southeastern Pacific, and South Atlantic Oceans, the reconstructed field reverts to the first-guess field, which shows different geographical patterns than the true patterns. This is because the localization strategy prevents the propagation of information from the observed regions to the data-sparse regions with a spatial distance larger than 20°. This illustrates a key assumption of the proposed method: the final analyzed field will revert to the first guess when very few observations are available (e.g., pre-1960). Under such circumstances, the current method trusts the first-guess field provided by the models, which leads to the pre-1960 error in Fig. 3b. The bias of the long-term trend due to this assumption was quantified and the results are presented in the following sections.

In 1961, the data coverage increased to greater than 10%, although the data were still distributed around the major cruise line routes. The geographical pattern of the reconstructed field agrees well with the truth field for 2014 (Fig. 4), and the global averaged error is lower compared with 1941 (Fig. 3b). This suggests that the unconstrained ensemble of model runs contains the correct type of variability to reconstruct the temperature field. Similar results can be found for the ocean sampling in 1981 and 2001.

In summary, we used the data from the last decade to determine the influencing radius and to perform a preliminary evaluation of our proposed method. We showed that the new mapping method is able to reconstruct the truth field post-1960 in an unbiased manner, but the reconstructed field shows a slight reversion to the first-guess field pre-1960. We note here that the principal shortcoming of this test is that data for this decade still does not have global coverage, especially in coastal areas and in the Southern Ocean.

5. Idealized tests using CMIP5 models

To evaluate the new method and examine potential improvement of the present analysis for historical ocean heat content change, a series of idealized tests were constructed by explicitly specifying a truth field. Pseudo observations were constructed by sampling the truth field according to the locations of historical observations from 1940 to 2014. Observational errors are added to the pseudo observations by generating random noise with zero mean and standard deviation calculated by Eq. (5). Then, the pseudo observations were analyzed by using a mapping method (discussed in the following). Finally, the mapped field we obtained was compared with the truth field to evaluate the performance of the reconstruction.

Three different mapping methods were applied separately for comparison (note that not all available mapping methods were compared):

1) The Gmean method uses an OHC calculated from the available data to represent the global OHC, potentially assuming that the OHC in data gaps is equal to the mean OHC in the available-data region. This method has been widely used in previous studies such as in Lyman et al. (2010) and Palmer et al. (2007).
2) The parameterized (PARAM) method is group of methods that used a parameterized error covariance (according to the spatial distance between the analyzed grid and the observations) and climatology (zero anomaly) as the first-guess field. Several major data centers have employed a similar method (e.g., Good et al. 2013; Ishii et al. 2003; Levitus et al. 2012). In this study, we adapted the method presented in Levitus et al. (2012).
3) The proposed method in this study (referred to as EnOI-DE/CMIP5 hereafter).

To test the performance of the three methods outlined above, and the following two experiments were constructed by setting two different truth fields:

1) The "truth ensemble" is the ensemble mean of the CMIP5 model outputs was used as the truth field. Because the first-guess field is the CMIP5 ensemble mean, this experiment showed the performance of the mapping methods when the first-guess field is exactly known.
2) The "truth select" is when one of the model outputs was randomly selected as the truth field. This experiment indicated the impact of a slightly biased first-guess field because the model selected has a different mean than the ensemble mean of the CMIP5 models. We used the Hadley Centre Coupled Model, version 3, as an example for better illustration, but the use of the other models would have led to the same conclusion.

The monthly and global means of 10-m (representative of SST) and 0–700-m-averaged temperature anomalies (representative of OHC) from 1940 to 2014
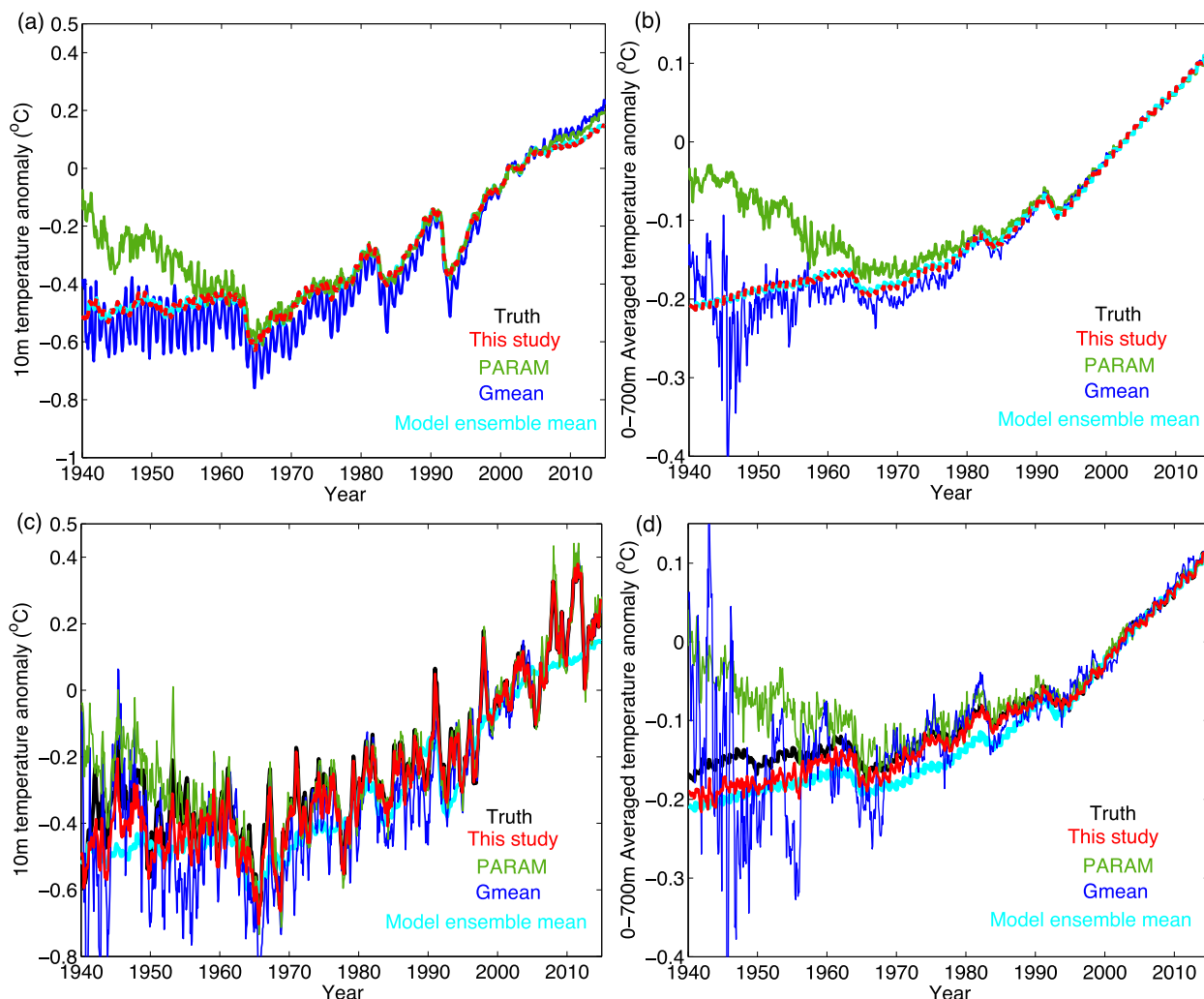
FIG. 5. Monthly time series of (a),(c) 10-m and (b),(d) 0–700-m-averaged temperature anomaly from 1940 to 2014 for two experiments: (top) the truth ensemble experiment and (bottom) the truth select experiment. The three mapping strategies are shown: the proposed method in red, the PARAM method in green, and the Gmean method in blue. The truth is shown in black, and the CMIP5 model ensemble mean is shown in cyan.

when the first-guess field was exact (the truth ensemble experiment) are presented in Fig. 5. Both the SST and OHC by the EnOI-DE/CMIP5 mapping exactly match the truth field, and the root-mean-square error (RMSE) is close to zero (Fig. 6), indicating that the new method is capable of reconstructing the historical ocean change when the first-guess and error covariance fields are well represented by the CMIP5 models. However, the historical SST and OHC time series from the PARAM mapping strongly revert to zero at earlier times, especially prior to 1970, mainly because large data gaps are present in the early era, which have been mostly filled by the first-guess field (zero anomaly). In addition, the Gmean approach results in a larger increase in the SST and OHC because the observations prior to 2000 were

mainly from the Northern Hemisphere, which had experienced stronger warming than the Southern Hemisphere (Cheng and Zhu 2014a).

When the first-guess field was substantially biased (i.e., the truth select experiment), the proposed mapping method also reconstructed the truth of the SST and 0–700-m-averaged temperature anomaly well, as shown in Fig. 5, especially after 1955. A significant improvement was found when using the new method over both the Gmean-mapping and PARAM-mapping methods because the new method resulted in a smaller mean error and RMSE, as shown in Fig. 6. However, pre-1955, the analyzed field slightly shifted to the CMIP5 ensemble mean. This confirms the analyses in the previous section, which indicate that the new method confers some trust
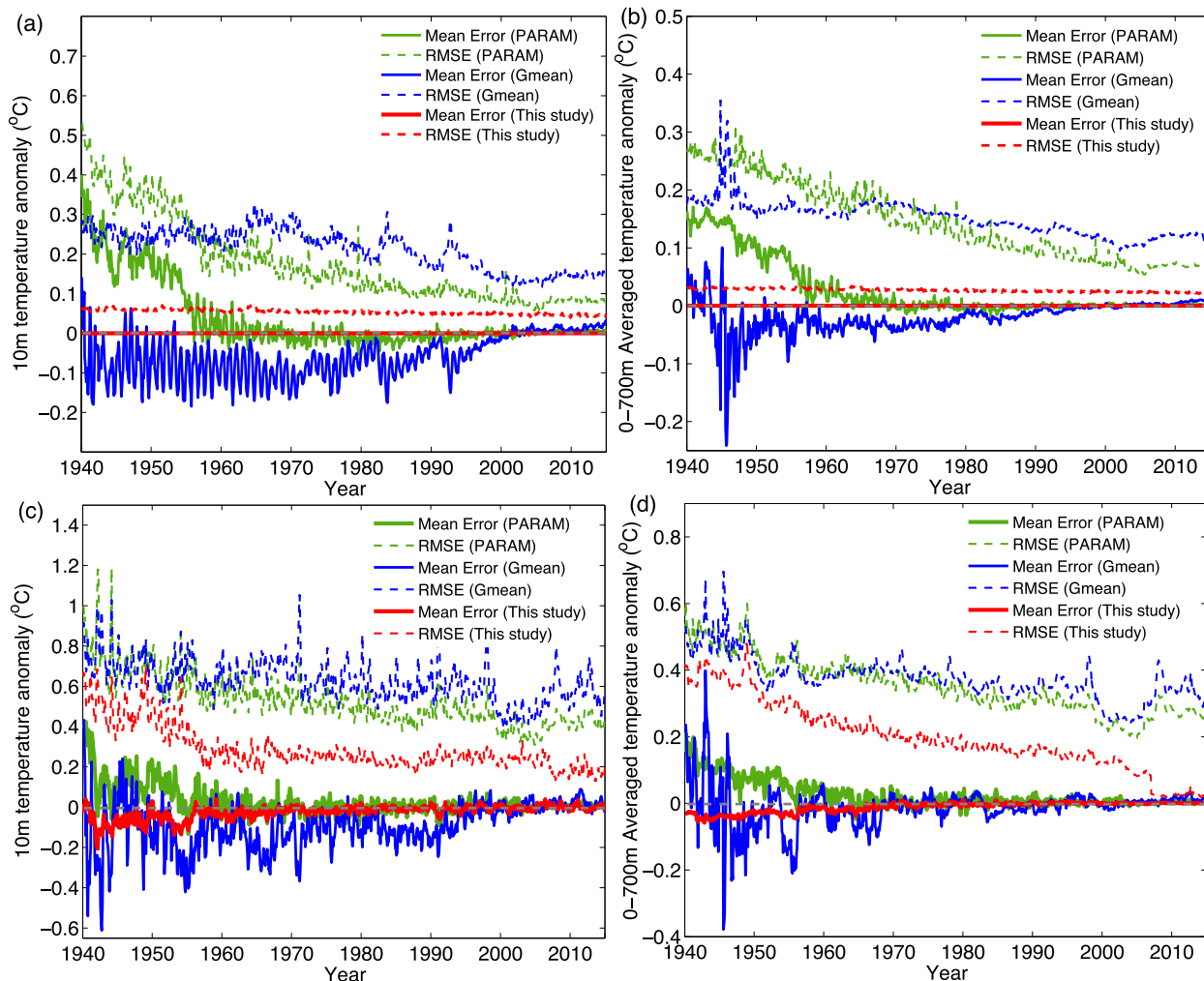
FIG. 6. Mean (solid curves) and RMSE (dashed curves) of the temperature difference between the analysis field and the truth field for (a),(b) truth ensemble experiments and (c),(d) truth select experiments for (left) the 10-m temperature anomaly and (right) the 0–700-m average temperature anomaly. The three mapping strategies are shown: the proposed method in red, the PARAM method in green, and the Gmean method in blue.

on the first-guess field when the data coverage is extremely sparse. Therefore, the performance of the new method pre-1960 might depend partly on how accurately the CMIP5 model ensemble represented the long-term temperature change. This error could either under- or overestimate the long-term trends of the historical ocean subsurface temperature change.

We included an additional experiment to quantify the impact of the first-guess field on the estimate of long-term trend of the SST and the 0–700-m temperature change based on the truth select framework. We randomly selected 10 models as the truth and then computed the long-term trend for the first guess (CMIP5 ensemble mean), truth, and the reconstructed field. Figure 7 shows the percent error of the long-term trend for both the first-guess and the reconstructed field

compared with the truth. Generally, the use of the proposed method could reduce the trend error to less than 20% from 1940 to 2014, 10% from 1970 to 2014, and 5% from 1990 to 2014, even if the trend error of the first-guess fields ranged from −60% to 80%. These results indicated that the new technique could capture long-term trends that were significantly different than the first guess. A worse first-guess field (with larger error) could lead to larger trend errors for the reconstructed field. However, according to Cheng et al. (2015b), the OHC trend of the CMIP5 ensemble mean over the 1970–2005 period was only 12% weaker than their observation-based estimate, suggesting that the multimodel ensemble mean was a reasonable choice for the first-guess field (the trend error will be less than 10%). Furthermore, a slightly smaller error was found for the trend at 10 m
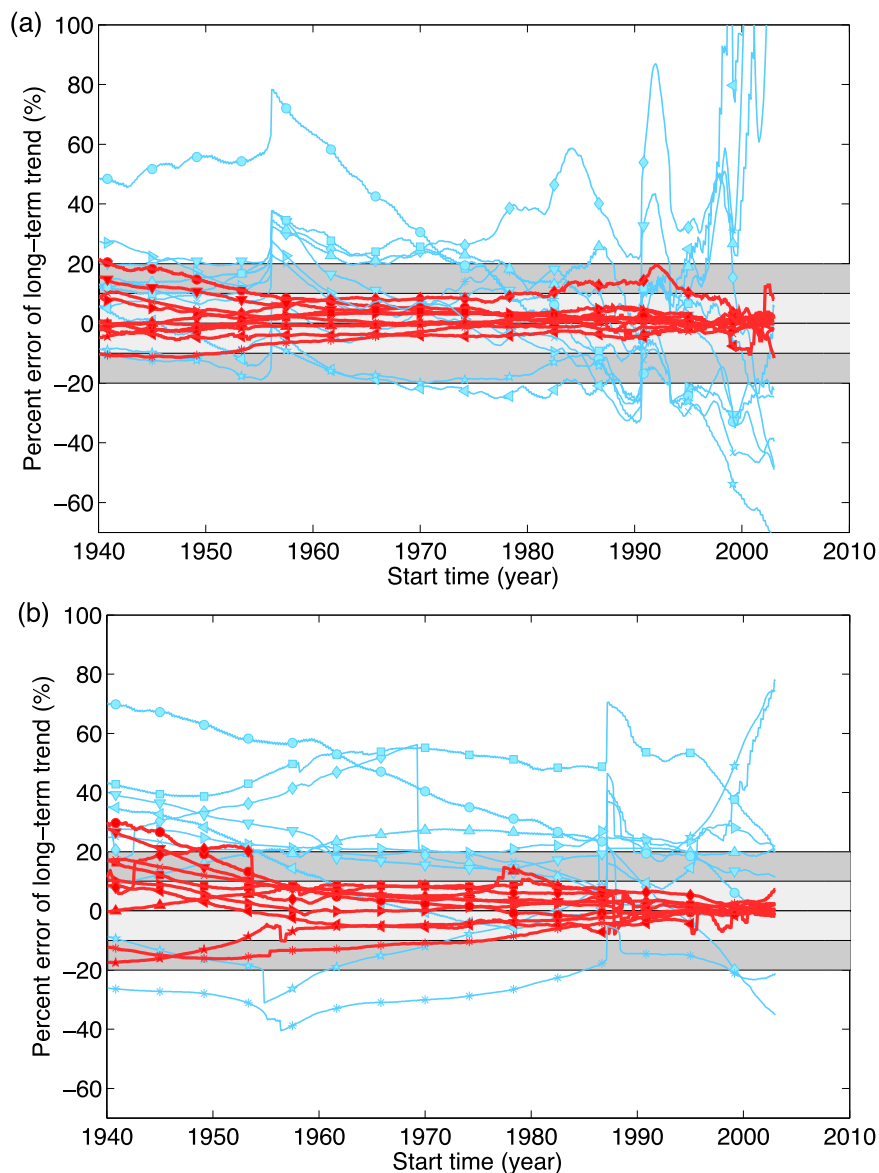
FIG. 7. Percent error of the long-term trend of the reconstructed field and the first-guess field compared to the truth field. The long-term trend was calculated from a start time (shown at time N in x axis) to 2014. The red curves show the percent error for reconstructed field, which was calculated by [trend(reconstructed) − trend(truth)]/[trend(truth)], and the light blue curves are the error for the first guess, [trend(first guess) − trend(truth)]/[trend(truth)], for (a) global averaged temperature trend at 10-m and (b) 0–700-m-averaged temperature. The dark gray shading shows the error from −20% to 20%, and the light gray shading is the error from −10% to 10%.

than the trend at 0–700 m, especially before 1960 because of better data coverage in the upper ocean than in the deeper ocean.

6. Assessment against withheld observations

The subsample tests and idealized tests indicated the potential for the improvement of the reconstruction of

the temperature fields from pseudo observations by using the EnOI-DE/CMIP5 method. In this section, we further evaluated the proposed approach by applying the mapping method to a portion of observations (80%) and using the remaining data (20%) for validation. To perform the subdivision, we randomly divided the observation dataset in each month into five subsets, each containing 20% of the data. Five different tests were
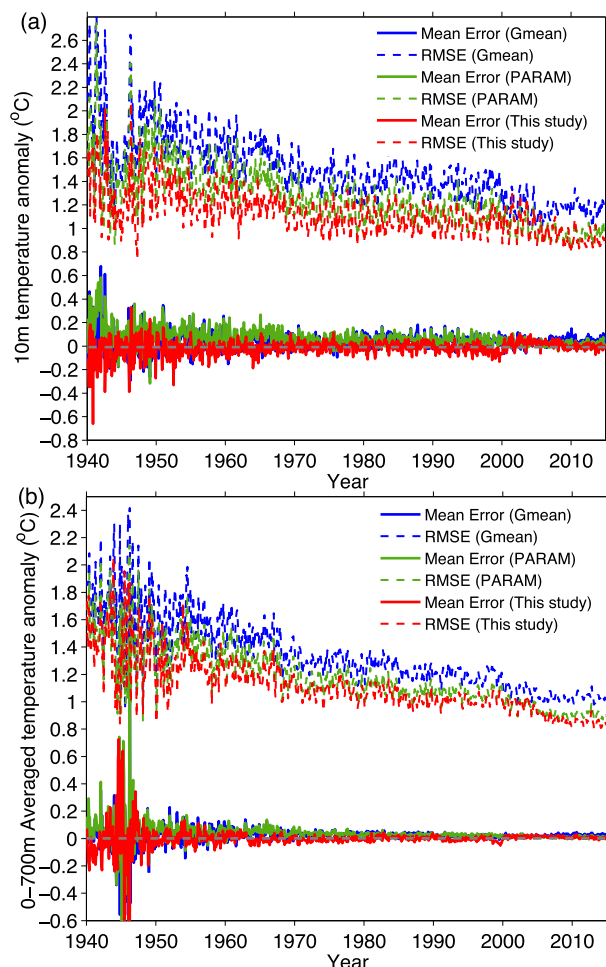
FIG. 8. Mean (solid curves) and RMSE (dashed curves) for the temperature difference between the analysis data and the withdrawn data (a) for the 10-m temperature anomaly and (b) for 0–700-m-averaged temperature anomaly.

performed, and each test used one of the subsets for validation and the remaining four for reconstruction. In this way, when performing five tests, all of the observations could be validated.

Figure 8 presents the monthly time series of the mean error and RMSE between the analyzed fields with respect to the withdrawn data (an average of the five tests). Based on the global average, the new method again resulted in a smaller error compared with the other two methods and reduced the RMSE by 0.24°C (0.38°C) compared with Gmean mapping and by 0.16°C (0.14°C) compared with the PARAM method at 0–700 m (at 10-m depth).

More detailed information was obtained by evaluating the geographical distribution of the mean error and the RMSE between the analyzed field and the withdrawn data when comparing the following three mapping methods: EnOI-DE/CMIP5 (Figs. 9a,b), Gmean

(Figs. 9c,d), and PARAM (Figs. 9e,f). Generally, all of the methods led to smaller mean errors and RMSE in low latitudes within 30°S–30°N than in higher latitudes and larger errors near the boundary current system because larger spatial and temporal mesoscale variabilities exist in the boundary current system, which are difficult to reconstruct. However, EnOI-DE/CMIP5 mapping resulted in much smaller errors than the other two methods over the global ocean. Compared with the PARAM method, better reconstruction of the temperature field was found for the Kuroshio and Gulf Stream with the proposed method. Those regions have been well sampled by historical ship-based observation systems, so this improvement indicates the effect of the covariance field. A globally consistent covariance (as in PARAM), which assumes the same spatial correlation over the global ocean, might not be accurate enough to represent the anisotropy of the covariance in those regions. By contrast, the models have some capability to simulate the general ocean circulation and could provide a better representation of the covariance for the boundary current system. In the same way, the proposed method also showed a smaller error than PARAM in the ACC regions. Furthermore, the Gmean method resulted in much larger errors in middle and high latitudes, indicating a key limitation of the traditional representative method. Because different temporal variations are present at different locations, filling the data gaps with the mean OHC in well-sampled regions led to the largest errors.

The results of the data withdrawal test again confirmed that the proposed method could reconstruct the historical temperature change. However, this test also is limited because the withdrawn data are mostly from the well-sampled regions. This indicates that this test mainly revealed the performance of the mapping methods in the data-rich regions.

7. Assessment with individual datasets

After the three different evaluations shown in the previous sections, the improvement of the proposed mapping method for reconstructing subsurface temperature fields is evident. Therefore, we applied the approach to the entire dataset and obtained the analyzed field of the ocean subsurface temperature. In this section, additional assessments were made by focusing on several different key metrics of the climate system because it is also important to investigate whether the new dataset is capable of representing the mean state and historical ocean variations on various time scales, such as a long-term trend, interannual variability, and decadal variability. We note that the assessments are
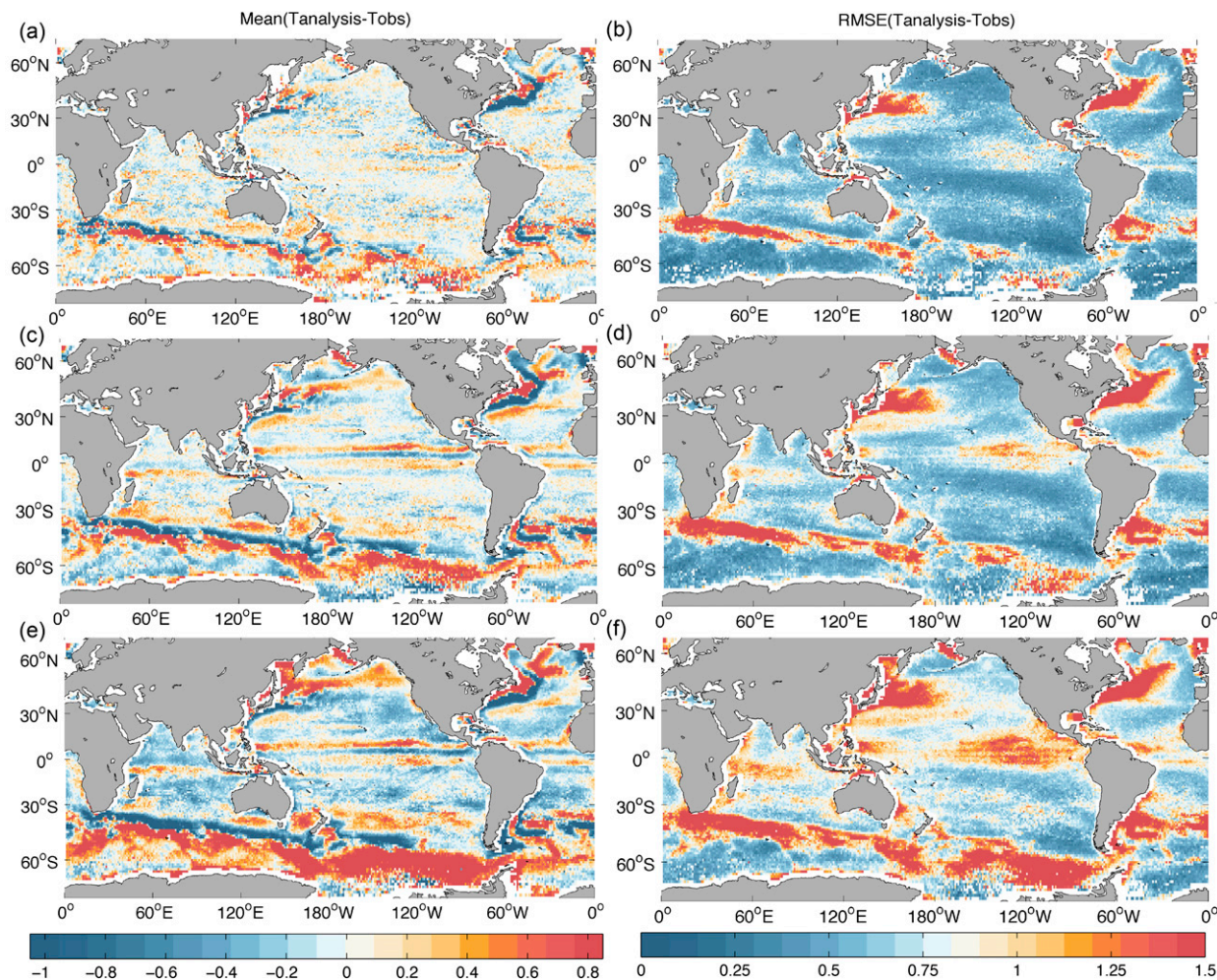
FIG. 9. Geographical distribution of the 0–700-m average temperature error (°C) and RMSE (°C) between the analysis field and withdrawn observations (average of the five withdrawn data tests, each withdrew 20% of the data) for (a),(b) the new method, EnOI-DE/CMIP5; (c),(d) PARAM mapping; and (e),(f) Gmean mapping.

mainly made for the sea surface temperature field because no independent subsurface dataset exists. However, these assessments also have important implications for the performance of the reconstruction of the ocean subsurface.

a. Reconstruction of the ocean climatology

Although the ultimate goal of this study is to provide a complete subsurface temperature anomaly field for climate change studies, it is also important to examine the temperature filed in the first evaluation. The 0–700-m average temperature during 2005–12, based on the fields reconstructed in this study, is presented in Fig. 10 and compared with the climatologies in *World Ocean Atlas 2013* (*WOA13*; Locarnini et al. 2013). *WOA13* is a well-developed and widely used ocean climatology that uses an objective analysis. The

most important development was its increased spatial resolution; two resolutions are presented ($1° \times 1°$ and $1/4° \times 1/4°$). Our 0–700-m average temperature climatology shows nearly an identical pattern with *WOA13* in the $1° \times 1°$ resolution, indicating a warm pool at the western boundary within 10°–30°N and 10°–30°S in each ocean basin. Although the 0–700-m average temperatures show small seasonal variability because of the large heat capacity of the ocean, ocean warming in the Northern (Southern) Hemisphere is still detectable in summer (winter). Furthermore, the *WOA13* climatology in the $1/4° \times 1/4°$ resolution reveals a similar pattern as the two climatologies discussed above but shows more mesoscale signals. This implies that this study could be potentially improved in the future by increasing the resolution to gain a better representation of the mesoscale eddies.
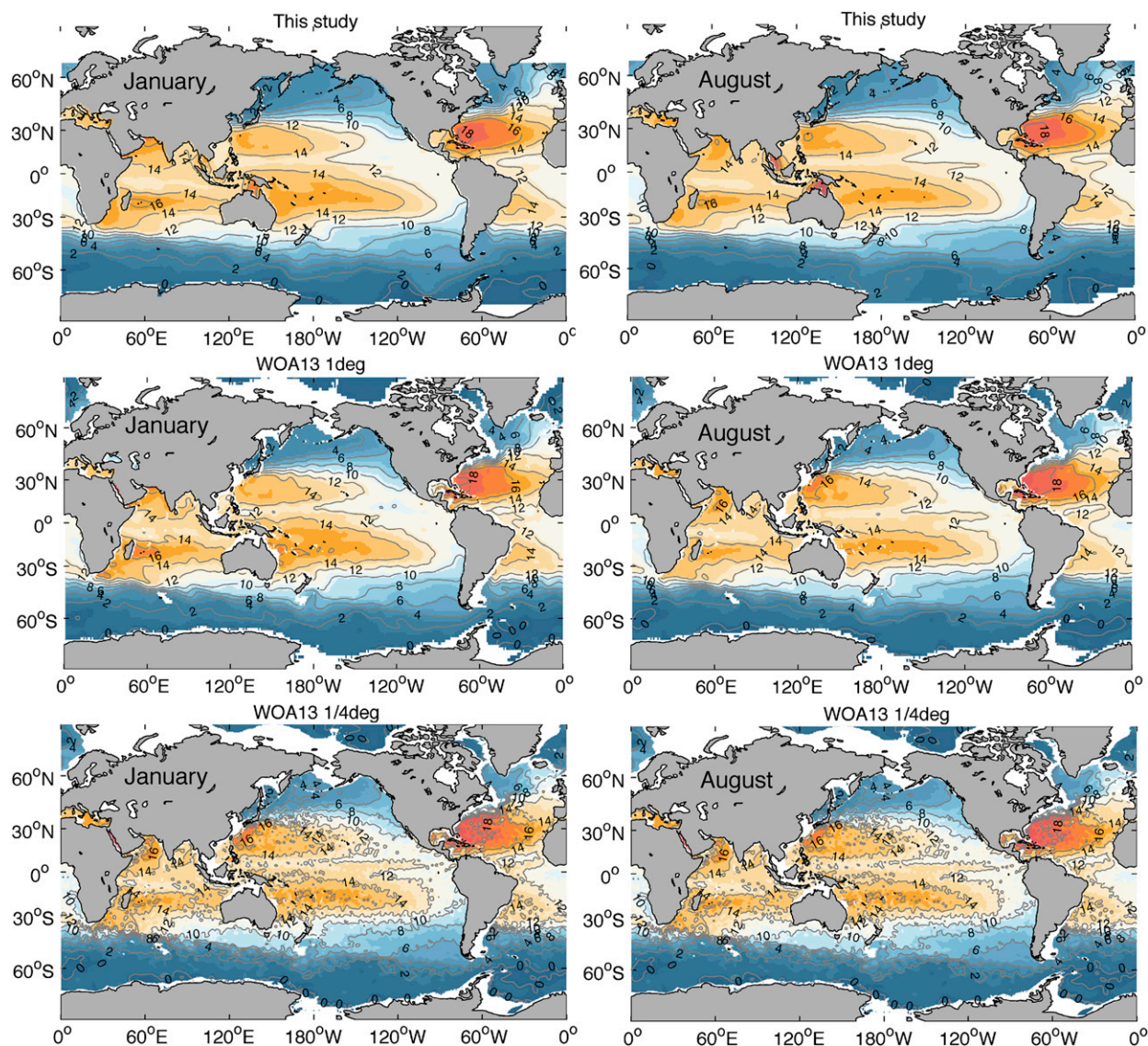
FIG. 10. Comparison of the 0–700-m ocean average temperature (°C) for 2005–12 between the reconstructed field from (top) this study and two versions of *WOA13*, (middle) 1° × 1° and (bottom) ¼° × ¼° resolution, shown separately for (left) January and (right) August.

b. Reconstruction of SST change

The sea surface temperature (SST) is an important variable associated with climate change, which has been carefully analyzed in various data centers (Kennedy et al. 2011; Rayner et al. 2003; Cowtan and Way 2014). SST variability at different time scales is well documented in the literature because it plays a crucial role in the air–sea interaction and dominates climate feedback related to global warming (i.e., the water cycle). Furthermore, it is also independent because it is measured by a different observation system (ship hulls, buoys, and insulated buckets) rather than via Argo or XBT devices. Therefore, the SST data can be considered an independent dataset for the

validation of our new dataset. In this section, we compared our analyzed field at a depth of 10 m with an independent and fully evaluated dataset, HadSST3 from the Met Office (Kennedy et al. 2011). Figure 11 shows the monthly time series of the global SST change from 1940 to 2014 in the reconstructed field compared with HadSST3. It is apparent that the interannual variability of the SST is nearly identical in both datasets, suggesting that the SST field has been well represented by the new approach. Regarding long-term changes, the EnOI-DE/CMIP5 method results in a linear trend of approximately $0.0069°C \, yr^{-1}$ from 1940 to 2014 and approximately $0.0127°C \, yr^{-1}$ from 1970 to 2014, showing a slightly stronger trend than the Hadley Centre SST, version 3 (HadSST3), dataset from 1940 to 2014
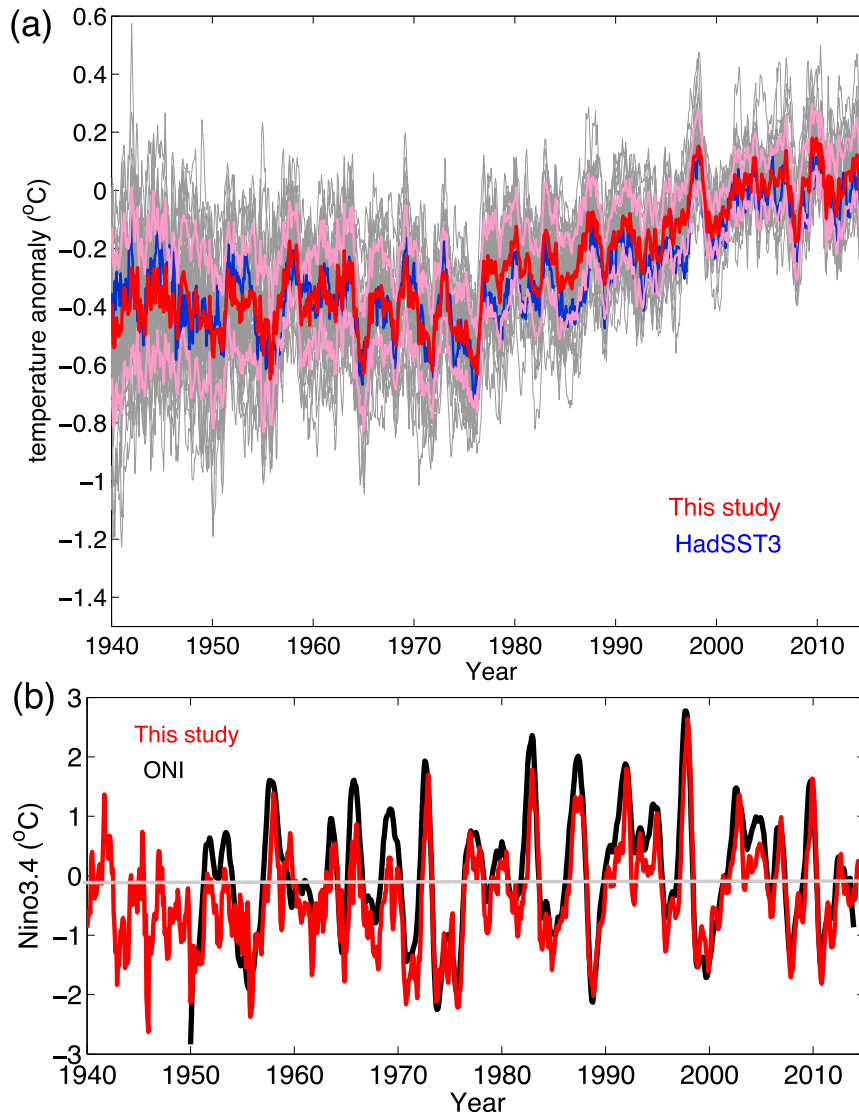
FIG. 11. (a) Global temperature anomaly time series at 10 m from 1940 to 2014: the ensemble mean is in red, and 40 (31 after 2006) ensemble members are in gray. Plus and minus one standard deviation is shown in pink. Global SST change in HadSST3 dataset is shown in blue. (b) Niño-3.4 index calculated by the proposed method is in red, and the ONI index is in black.

$(\sim 0.0057^{\circ}\text{C yr}^{-1})$ and a near-identical trend from 1970 to 2014 $(\sim 0.0128^{\circ}\text{C yr}^{-1})$. These differences can be attributed to 1) the difference in the land mask used by the two groups; 2) the difference in the reconstructed methods, in particular, the results of our method might be slightly reverted to the first-guess field pre-1960; and 3) uncertainties in both the sea surface temperature datasets (HadSST3) and the near-surface datasets (this study).

The correlation of SST change in each $1^{\circ} \times 1^{\circ}$ grid between the reconstructed dataset and the HadISST dataset (Rayner et al. 2003) is shown in Fig. 12, where the long-term linear trend and the seasonal cycle in each

grid has been removed. The results showed a significant positive correlation within 30°S–60°N, as follows: $R > 0.5$ in general from 1940 to 2014 and $R > 0.7$ from 1970 to 2014. However, a small or even a negative correlation in the Southern Ocean (60°–30°S) was apparent, which was due to the lack of observations for both subsurface and sea surface temperatures (i.e., HadISST).

c. Reconstruction of interannual variability (ENSO)

ENSO is the most important signal on the interannual scale in the climate system; it is able to affect weather conditions over the entire globe by teleconnection
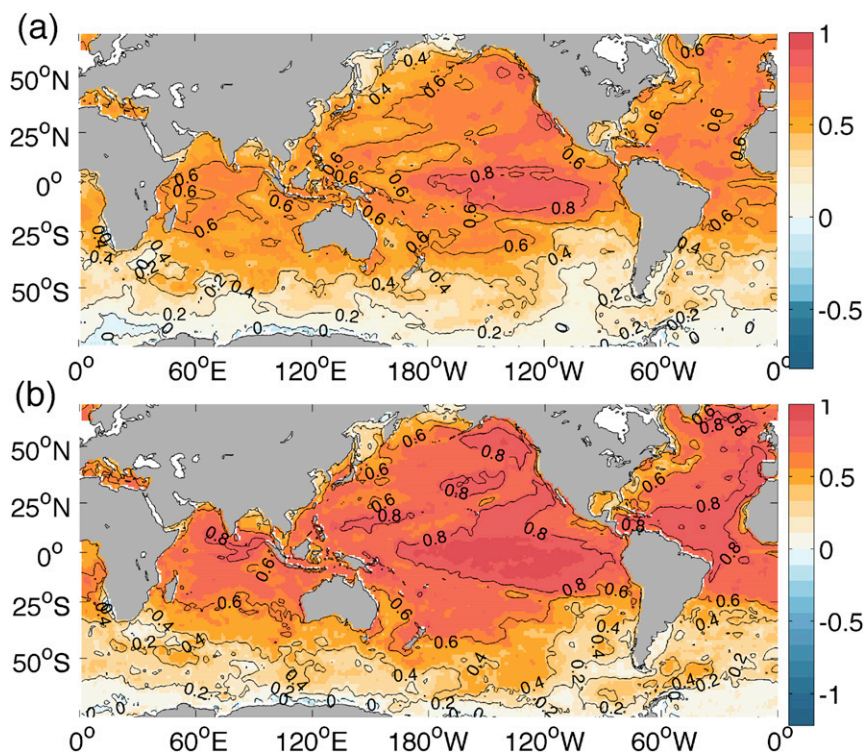
FIG. 12. Correlation of temperature change between HadISST and the reconstructed 10-m ocean temperature field in this study. The correlation is calculated in each $1° \times 1°$ grid (a) from 1940 to 2014 and (b) from 1970 to 2014.

mechanisms. As indicated by Palmer and McNeall (2014) and Taylor et al. (2012), the timing of an unforced variability such as ENSO only matches observations by coincidence. Thus, it remains to be observed whether the temperature field obtained from the CMIP5 data in the EnOI-DE/CMIP5 method represented the correct timing of an unforced variability such as ENSO. Here, ENSO will be indicated by the Niño-3.4 index.

The global SST time series presented in section 7b has already shown that the interannual variability in the reconstructed SST field agreed with the observations. Figure 11b shows a further calculation of the Niño-3.4 index by using the reconstructed field compared to the index obtained with the Niño-3.4 index provided by oceanic Niño index (ONI) from NOAA. The two indexes are almost identical ($R > 0.8$), indicating that the timing of the ENSO variability was well represented by the reconstructed field. However, we note here that the reconstruction based on temperature anomaly field after subtracting 12 monthly climatologies is difficult to un-biased catching ENSO, because ENSO is phase locking in season (always get to the peak in winter) and asymmetric between its positive and negative phase. This leads to the residuals of ENSO signals in the monthly climatology, which potentially bias the representation of

ENSO in the anomaly field. How well the reconstructed field in this study in representing ENSO compared with the existing methods requires a more comprehensive study in the future, which is out of scope of this study.

d. Reconstruction of decadal variability (PDO)

Decadal variability is another important feature of the climate system. The interdecadal Pacific oscillation (IPO) is one of the predominant signals of climate variability on a decadal scale (Zhang et al. 1997), which is responsible for the recent "climate hiatus" that has been documented in many recent studies (Meehl et al. 2011; Kosaka and Xie 2013; Meehl et al. 2013; Trenberth and Fasullo 2013; England et al. 2014; Trenberth et al. 2014b; Cheng et al. 2015a; Trenberth 2015). Therefore, we calculated the first empirical orthogonal function (EOF) mode of the Pacific SST field within 30°S–45°N. The spatial pattern of the first EOF mode is shown in Fig. 13a and displays a horseshoe pattern that includes a maximum positive signal in the eastern Pacific Ocean in low latitudes and two negative centers in middle latitudes in both hemispheres. This pattern is well documented in the literature and reveals the low-frequency variability of the air–sea interaction mode in the Pacific Ocean. The time series of this mode are presented in Fig. 13b,
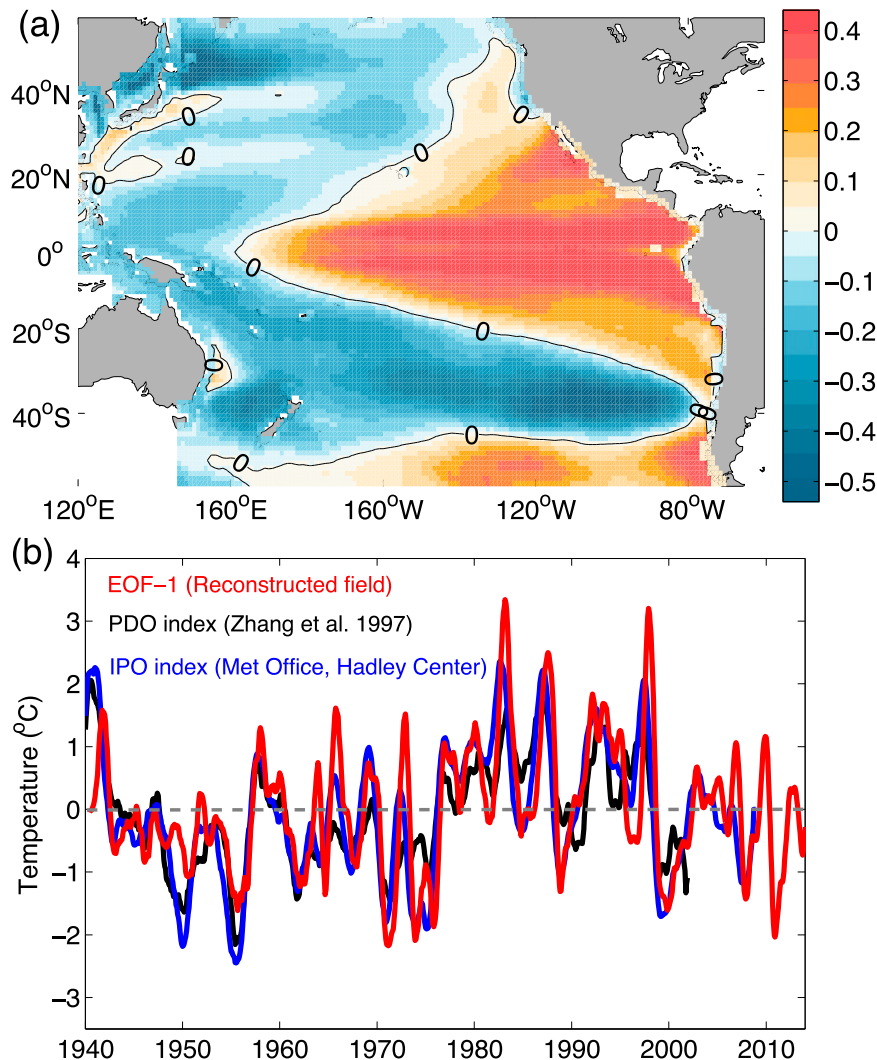
FIG. 13. (a) Spatial pattern and (b) time series of the first EOF mode (in red) of the reconstructed temperature field at 10 m in Pacific Ocean from 55°S to 55°N. The time series of the first EOF mode is compared with PDO index shown in Zhang et al. (1997) and IPO index provided by the Met Office Hadley Centre (Folland et al. 2002).

appearing as a negative phase during the 1950–75 and 1998–2014 periods and as a positive phase during the 1976–97 period. For comparison, the IPO index provided by the Met Office Hadley Centre (Folland et al. 2002) and the Pacific decadal oscillation (PDO) index provided by Zhang et al. (1997) are also shown in Fig. 13b. The three time series are consistent with each other, indicating that the IPO (or PDO) signal can be reconstructed well by the new approach.

8. The reconstructed field and its uncertainties

Based on the evaluation as outlined in the previous sections, we found that the EnOI-DE/CMIP5 mapping method showed robust improvement in the reconstruction of the ocean surface and subsurface temperature fields at different time scales. As indicated in section 3, the uncertainties of the reconstructed field can also be assessed, as represented by updating the members of the ensemble.

For example, Fig. 11 displays the SST time series of the ensemble members with their standard deviation in pink. Larger errors are apparent at earlier times: ±0.2°C in 1960 compared with ±0.05°C in 2014. The reduction of such errors might be tied to both an increase in data coverage and the improvement of the instrument quality of the observations. Moreover, the uncertainty was smaller during 1999–2005 than during the other periods because the 1999–2005 climatology was used. The global averaged anomalies for the different models were all
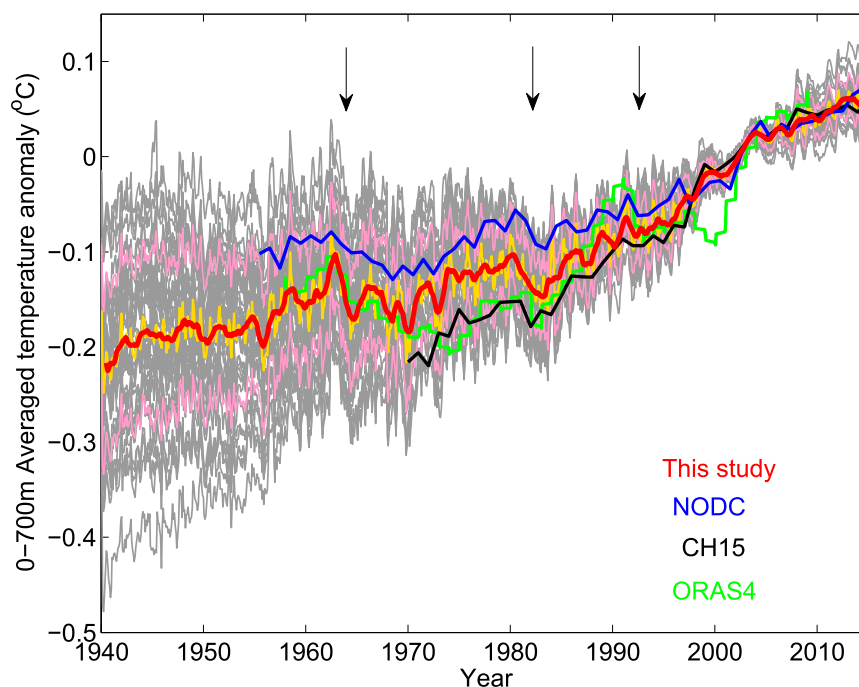
FIG. 14. Historical upper 0–700-m ocean heat content change (in values of 0–700-m-averaged temperature anomaly) using the proposed method: ensemble mean is shown in red and 40 ensemble members (31 after 2006) in gray. Plus and minus one standard deviation is shown in pink. The yellow curve shows the monthly mean, and the red curve is the smoothed time series by a 12-point running smoother. The new estimate is compared with three independent estimates of Levitus et al. (2012) in blue (NODC), Cheng et al. (2015b) in black (CH15), and Balmaseda et al. (2013) in green (ORAS4). The black arrows denote the volcano eruptions: Mount Agung in 1963, El Chichón in March–April 1982, and Mount Pinatubo in June 1991. All of the time series are referenced to 2005.

approximately zero during this period, which led to a smaller analysis error for the globally averaged metrics such as the global SST.

Similarly, the 0–700-m-averaged temperature change (representative of the OHC) was calculated from 1940 to 2014 using the proposed method together with the 40 (31 after 2006) ensemble members (gray curves) and the standard deviation (pink curves) as shown in Fig. 14. This estimate can be regarded as another independent OHC estimate compared with the existing estimates presented in Ishii et al. (2003), Willis et al. (2004), Smith and Murphy (2007), Domingues et al. (2008), Palmer and Haines (2009), Roemmich and Gilson (2009), Levitus et al. (2012), Good et al. (2013), and Cheng et al. (2015b). The new EnOI-DE/CMIP5 method resulted in a long-term linear trend of $0.0034 \pm 0.0010°C\,yr^{-1}$ ($0.20 \pm 0.07\,W\,m^{-2}$ averaged over the global surface) from 1940 to 2014 and $0.0052 \pm 0.0009°C\,yr^{-1}$ ($0.31 \pm 0.06\,W\,m^{-2}$) from 1970 to 2014. The three other independent estimates (Cheng et al. 2015b; Levitus et al. 2012; Balmaseda et al. 2013) are also shown in Fig. 14. The new estimate from this study showed a slightly

weaker OHC trend compared with Cheng et al. (2015b; $\sim0.35 \pm 0.12\,W\,m^{-2}$) from 1970 to 2014 and Balmaseda et al. (2013; ~0.30–$0.34\,W\,m^{-2}$) from 1970 to 2009. However, it showed a much stronger trend from 1970 to 2014 than Levitus et al. (2012; $\sim0.24\,W\,m^{-2}$). As indicated by the tests in this study, the traditional parameterized methods (i.e., Ishii et al. 2003; Smith and Murphy 2007; Levitus et al. 2012) most likely underestimate the long-term trend because of the choice of the first-guess field.

Moreover, the estimates by Cheng et al. (2015b), NOAA/NODC, and this study show a very similar interannual oscillation, although the magnitude is not consistent. The interannual variation of the upper OHC is linked to the ENSO variation as investigated in Roemmich and Gilson (2011) by using Argo data since 2004. Moreover, a strong OHC shift from 2001 to 2003 was shown in Levitus et al. (2012), to be partly (nearly by half) spurious, which was caused by the change in the observation system from a ship-based system to the Argo system (Cheng and Zhu 2014a). As indicated in Fig. 14, the net temperature change within 2001–03 is

lower by half in both Cheng et al. (2015b) and in this study, compared with Levitus et al. (2012). Furthermore, the OHC changes forced by three major volcano eruptions are apparent in the reconstructed field: Mount Agung in 1963, El Chichón in March–April 1982, and Mount Pinatubo in June 1991.

The calculation of the geographical distribution of the long-term change in the ocean heat content is also valuable. Figure 15 shows the linear trend of the temperature change at 10 m from 1970 to 2014 compared with that based on HadISST, and the zonal average temperature change at each depth from 1 to 700 m. The sea surface temperature field shows large-scale warming over the global ocean, indicating signals of anthropogenic forcing. More rapid warming of the poleward flowing western boundary currents such as the East Australian Current and the Kuroshio can be detected (Wu et al. 2012).

The zonal trend of temperature change shown in Fig. 15c reveals the modification of the ocean circulation and the downward transportation of heat. The large-scale warming from 10° to 40°N and from 20° to 50°S indicates the downward transport of warming by subtropical cells (Levitus et al. 2009; Rhein et al. 2013) and is consistent with poleward displacement of the mean temperature field. The stronger warming within 50°–40°S might be linked to the poleward displacement of the ACC and the southern limb of the subtropical gyres, as illustrated by Gille (2008) and Alory et al. (2007). The weak warming (or slight cooling) at depths between 20°S and 10°N has also been shown by previous studies (i.e., Rhein et al. 2013) and is consistent with a southward shift of cooler water near the equator. This is probably because of the strengthening of the shallow subtropical cell due to the intensifying tropical trade winds since 1990s (England et al. 2014). The cooling of the Southern Ocean is apparent at the upper 100 m, consistent with the recent increase of sea ice coverage.

9. Concluding remarks

In the latest IPCC report (Rhein et al. 2013), the available upper-ocean heat content estimates from 1970 to 2010 show large differences ranging from 74 TW (Smith and Murphy 2007) to 137 TW (Domingues et al. 2008), implying a large uncertainty of calculation. The community has made a great effort to reduce the errors in the OHC calculation as documented in the literature. Among these efforts are the removal of XBT biases (Abraham et al. 2014; Cowley et al. 2013; Levitus et al. 2009; Schwalbach et al. 2014; Shepard et al. 2014; Cheng et al. 2016), the selection of an appropriate climatology (Lyman and Johnson 2014;

Cheng and Zhu 2015), the removal of errors due to the insufficient vertical resolution of the data (Cheng and Zhu 2014b), and the proposal of new mapping methods (Lyman et al. 2010; Levitus et al. 2012). In this study, we have focused on a new mapping strategy motivated by the deficiencies in some of the available methods: 1) the choice of climatology (zero anomaly) as the first-guess leads to an overreversion of the OHC estimate toward zero, and 2) the parameterized error covariance is empirical and might not work well for some regions such as the boundary current regions. Based on this motivation, our new strategy uses an ensemble of CMIP5 models to provide the first-guess field and error covariance and shows better performance in reconstructing the historical ocean subsurface temperature field.

However, we do not claim that this method is the best because it is still unclear whether it can perfectly fill the data gaps in the historical dataset. The difficulty is that no independent data for validation exist, especially for the ocean subsurface. A potential strategy to fully evaluate this method is the use of several different synthetic datasets, such as high-resolution models, sea level data, etc., which requires further effort. Furthermore, comprehensive and international collaborations are required in the future to fully understand the performance of different mapping methods.

It is worth noting here that a potential problem with the new method is that the temperature field (and global OHC) from the earlier eras (i.e., prior to 1960) based on the EnOI-DE/CMIP5 method might slightly revert to the ensemble mean of the CMIP5 model outputs. Therefore, the accuracy of this method might depend on the performance of the CMIP5 models, and future improvements of the models will positively feed back to the reconstruction of historical subsurface temperature field.

It is also noteworthy that further applications of the proposed strategy include the reconstruction of the deeper ocean temperature field below 700 m because increasingly more evidence suggests the importance of the deeper ocean in Earth's energy budget (Balmaseda et al. 2013; Palmer et al. 2011; Purkey and Johnson 2010; Trenberth et al. 2014a). However, the difficulty is that deeper ocean measurements are even sparser than those for the upper ocean because most of the XBT instruments have a maximum depth of 700–750 m. Another possible application is the reconstruction of the ocean subsurface salinity field, which plays an important role in monitoring the sea level rise (Church and White 2011; Church et al. 2011; Durack et al. 2014b) and the water cycle (Durack et al. 2012).
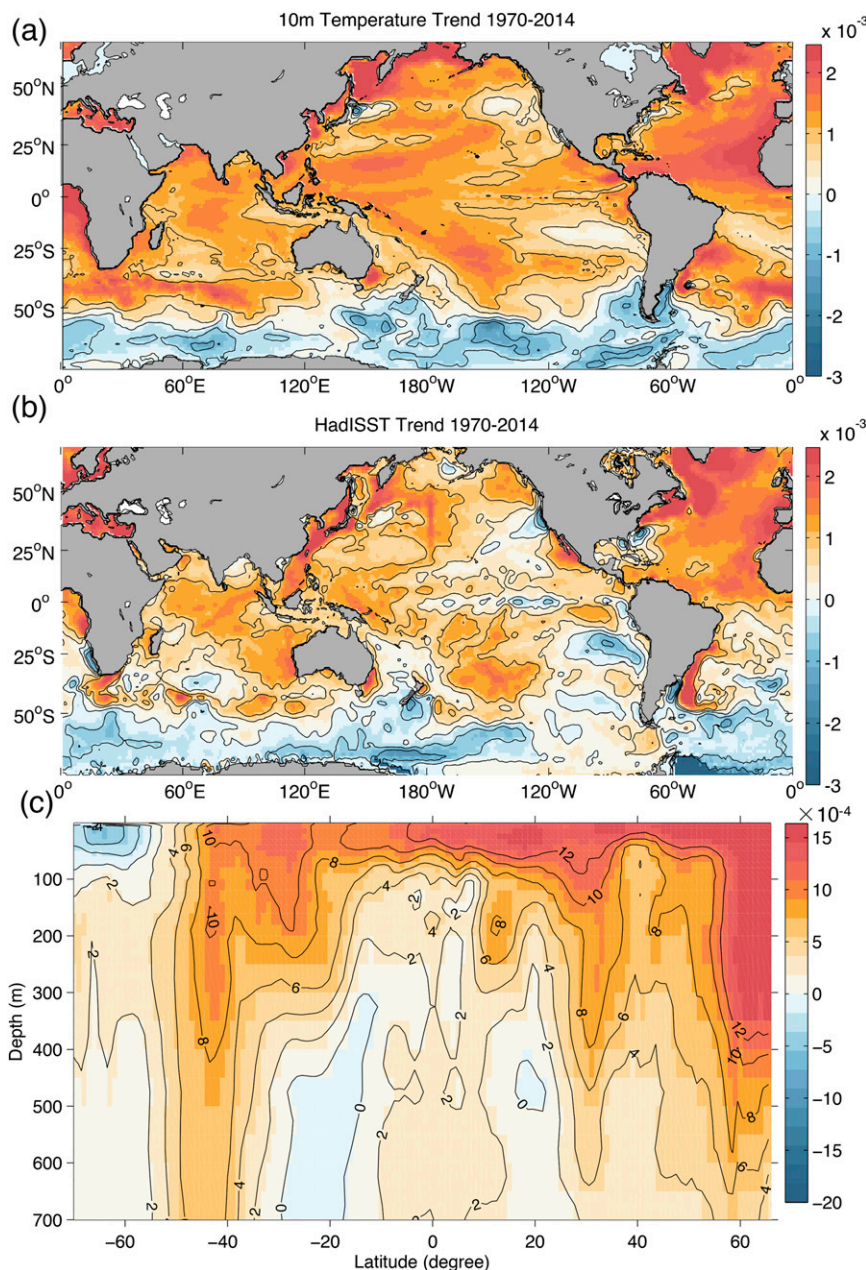
FIG. 15. (a) Linear trend of the temperature change at 10 m from 1970 to 2014 in the reconstructed dataset in this study. (b) As in (a), but for the HadISST dataset. (c) Zonal-averaged temperature trend as a function of latitude and depth from 1970 to 2014.

REFERENCES

Abraham, J. P., and Coauthors, 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483, doi:10.1002/rog.20022.

——, and Coauthors, 2014: Modeling and numerical simulation of the forces acting on a sphere during early-water entry. *Ocean Eng.*, **76**, 1–9, doi:10.1016/j.oceaneng.2013.11.015.

AchutaRao, K. M., and Coauthors, 2007: Simulated and observed variability in ocean temperature and heat content. *Proc. Natl. Acad. Sci. USA*, **104**, 10 768–10 773, doi:10.1073/pnas.0611375104.

Allan, R. P., C. Liu, N. G. Loeb, M. D. Palmer, M. Roberts, D. Smith, and P. L. Vidale, 2014: Changes in global net radiative imbalance 1985–2012. *Geophys. Res. Lett.*, **41**, 5588–5597, doi:10.1002/2014GL060962.

Alory, G., S. Wijffels, and G. Meyers, 2007: Observed temperature trends in the Indian Ocean over 1960–1999 and associated mechanisms. *Geophys. Res. Lett.*, **34**, L02606, doi:10.1029/2006GL028044.

Balmaseda, M. A., K. E. Trenberth, and E. Källén, 2013: Distinctive climate signals in reanalysis of global ocean heat content. *Geophys. Res. Lett.*, **40**, 1754–1759, doi:10.1002/grl.50382.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Boyer, T. P., and Coauthors, 2013: *World Ocean Database 2013*. S. Levitus, Ed., NOAA Atlas NESDIS 72, 209 pp. [Available online at http://data.nodc.noaa.gov/woa/WOD13/DOC/wod13_intro.pdf.]

Carton, J. A., and B. S. Giese, 2008: A reanalysis of ocean climate using simple ocean data assimilation (SODA). *Mon. Wea. Rev.*, **136**, 2999–3017, doi:10.1175/2007MWR1978.1.

——, and A. Santorelli, 2008: Global decadal upper-ocean heat content as viewed in nine analyses. *J. Climate*, **21**, 6015–6035, doi:10.1175/2008JCLI2489.1.

Chang, Y.-S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2013: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dyn.*, **40**, 775–803, doi:10.1007/s00382-012-1412-2.

Cheng, L., and J. Zhu, 2014a: Artifacts in variations of ocean heat content induced by the observation system changes. *Geophys. Res. Lett.*, **41**, 7276–7283, doi:10.1002/2014GL061881.

——, and ——, 2014b: Uncertainties of the ocean heat content estimation induced by insufficient vertical resolution of historical ocean subsurface observations. *J. Atmos. Oceanic Technol.*, **31**, 1383–1396, doi:10.1175/JTECH-D-13-00220.1.

——, and ——, 2015: Influences of the choice of climatology on ocean heat content estimation. *J. Atmos. Oceanic Technol.*, **32**, 388–394, doi:10.1175/JTECH-D-14-00169.1.

——, ——, R. Cowley, T. Boyer, and S. Wijffels, 2014: Time, probe type and temperature variable bias corrections to historical expendable bathythermograph observations. *J. Atmos. Oceanic Technol.*, **31**, 1793–1825, doi:10.1175/JTECH-D-13-00197.1.

——, F. Zheng, and J. Zhu, 2015a: Distinctive ocean interior changes during the recent warming slowdown. *Sci. Rep.*, **5**, 14346, doi:10.1038/srep14346.

——, J. Zhu, and J. Abraham, 2015b: Global upper ocean heat content estimation: Recent progress and the remaining

challenges. *Atmos. Oceanic Sci. Lett.*, **8**, 333–338, doi:10.3878/AOSL20150031.

——, and Coauthors, 2016: XBT Science: Assessment of instrumental biases and errors. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-15-00031.1, in press.

Church, J. A., and N. J. White, 2011: Sea-level rise from the late 19th to the early 21st century. *Surv. Geophys.*, **32**, 585–602, doi:10.1007/s10712-011-9119-1.

——, and Coauthors, 2011: Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.*, **38**, L18601, doi:10.1029/2011GL048794.

Cowley, R., S. Wijffels, L. Cheng, T. Boyer, and S. Kizu, 2013: Biases in expendable bathythermograph data: A new view based on historical side-by-side comparisons. *J. Atmos. Oceanic Technol.*, **30**, 1195–1225, doi:10.1175/JTECH-D-12-00127.1.

Cowtan, K., and R. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, doi:10.1002/qj.2297.

Dominguez, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn, 2008: Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**, 1090–1093, doi:10.1038/nature07080.

Durack, P., S. E. Wijffels, and R. J. Matear, 2012: Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science*, **336**, 455–458, doi:10.1126/science.1212222.

——, P. J. Gleckler, F. Landerer, and K. E. Taylor, 2014a: Quantifying underestimates of long-term upper-ocean warming. *Nat. Climate Change*, **4**, 999–1005, doi:10.1038/nclimate2389.

——, S. Wijffels, and P. J. Gleckler, 2014b: Long-term sea-level change revisited: The role of salinity. *Environ. Res. Lett.*, **9**, 114017, doi:10.1088/1748-9326/9/11/114017.

England, H. M., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, doi:10.1038/nclimate2106.

Evensen, G., 1994: Sequential Data Assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:10.1029/94JC00572.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

——, 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.*, **54**, 539–560, doi:10.1007/s10236-004-0099-2.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Folland, C. K., J. A. Renwick, M. J. Salinger, and A. B. Mullan, 2002: Relative influences of the interdecadal Pacific oscillation and ENSO on the South Pacific convergence zone. *Geophys. Res. Lett.*, **29**, 1643, doi:10.1029/2001GL014201.

Freeland, H., and Coauthors, 2010: Argo—A decade of progress. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, Vol. 2, J. Hall et al., Eds., ESA Publication WPP-306, 14 pp., doi:10.5270/OceanObs09.cwp.32.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.

Gille, S. T., 2008: Decadal-scale temperature trends in the Southern Hemisphere ocean. *J. Climate*, **21**, 4749–4765, doi:10.1175/2008JCLI2131.1.

Gleckler, P. J., and Coauthors, 2012: Human-induced global ocean warming on multidecadal timescales. *Nat. Climate Change*, **2**, 524–529, doi:10.1038/nclimate1553.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, doi:10.1002/2013JC009067.

Ishii, M., and M. Kimoto, 2009: Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.*, **65**, 287–299, doi:10.1007/s10872-009-0027-7.

——, ——, and M. Kachi, 2003: Historical ocean subsurface temperature analysis with error estimates. *Mon. Wea. Rev.*, **131**, 51–73, doi:10.1175/1520-0493(2003)131<0051: HOSTAW>2.0.CO;2.

Kalman, R. E., 1960: A new approach to linear filter and prediction problems. *J. Basic Eng.*, **82**, 35–45, doi:10.1115/1.3662552.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties *J. Geophys. Res.*, **116**, D14103, doi:10.1029/2010JD015218.

Kosaka, Y., and S. P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, doi:10.1038/nature12534.

Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, doi:10.1029/2008GL037155.

——, and Coauthors, 2012: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.*, **39**, L10603, doi:10.1029/2012GL051106.

Locarnini, R. A., and Coauthors, 2013: *Temperature*. Vol. 1, *World Ocean Atlas 2013*, NOAA Atlas NESDIS 73, 40 pp. [Available online at http://data.nodc.noaa.gov/woa/WOA13/DOC/WOA13_vol1.pdf.]

Lyman, J. M., and G. C. Johnson, 2008: Estimating annual global upper-ocean heat content anomalies despite irregular in situ ocean sampling. *J. Climate*, **21**, 5629–5641, doi:10.1175/2008JCLI2259.1.

——, and ——, 2014: Estimating global ocean heat content changes in the upper 1800 m since 1950 and the influence of climatology choice. *J. Climate*, **27**, 1945–1957, doi:10.1175/JCLI-D-12-00752.1.

——, S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis, 2010: Robust warming of the global upper ocean. *Nature*, **465**, 334–337, doi:10.1038/nature09043.

Meehl, G. A., J. M. Arblaster, J. Y. Fasullo, A. Hu, and K. E. Trenberth, 2011: Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nat. Climate Change*, **1**, 360–364, doi:10.1038/nclimate1229.

——, A. Hu, J. M. Arblaster, J. Y. Fasullo, and K. E. Trenberth, 2013: Externally forced and internally generated decadal climate variability associated with the interdecadal Pacific oscillation. *J. Climate*, **26**, 7298–7310, doi:10.1175/JCLI-D-12-00548.1.

Palmer, M. D., and K. Haines, 2009: Estimating oceanic heat content change using isotherms. *J. Climate*, **22**, 4953–4969, doi:10.1175/2009JCLI2823.1.

——, and D. J. McNeall, 2014: Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environ. Res. Lett.*, **9**, 034016, doi:10.1088/1748-9326/9/3/034016.

——, K. Haines, S. F. B. Tett, and T. J. Ansell, 2007: Isolating the signal of ocean global warming. *Geophys. Res. Lett.*, **34**, L23610, doi:10.1029/2007GL031712.

——, D. McNeall, and N. Dunstone, 2011: Importance of the deep ocean for estimating decadal changes in Earth's radiation balance. *Geophys. Res. Lett.*, **38**, L13707, doi:10.1029/2011GL047835.

——, and Coauthors, 2016: Ocean heat content variability and change in an ensemble of ocean reanalyses. *Climate Dyn.*, doi:10.1007/s00382-015-2801-0, in press.

Purkey, S., and G. Johnson, 2010: Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *J. Climate*, **23**, 6336–6351, doi:10.1175/2010JCLI3682.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.

Rhein, M., and Coauthors, 2013: Observations: Ocean. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 255–315.

Riser, S. C., and Coauthors, 2016: Fifteen years of ocean observations with the global Argo array. *Nat. Climate Change*, **6**, 145–153, doi:10.1038/nclimate2872.

Roemmich, D., and J. Gilson, 2009: The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Prog. Oceanogr.*, **82**, 81–100, doi:10.1016/j.pocean.2009.03.004.

——, and ——, 2011: The global ocean imprint of ENSO. *Geophys. Res. Lett.*, **38**, L13606, doi:10.1029/2011GL047992.

——, W. J. Gould, and J. Gilson, 2012: 135 years of global ocean warming between the Challenger expedition and the Argo Programme. *Nat. Climate Change*, **2**, 425–428, doi:10.1038/nclimate1461.

——, J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. *Nat. Climate Change*, **5**, 240–245, doi:10.1038/nclimate2513.

Sakov, P., and P. R. Oke, 2008a: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus*, **60A**, 361–371, doi:10.1111/j.1600-0870.2007.00299.x.

——, and ——, 2008b: Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.*, **136**, 1042–1053, doi:10.1175/2007MWR2021.1.

Schwalbach, D. S., T. Shepard, S. Kane, D. Siglin, T. Harrington, and J. P. Abraham, 2014: Effect of impact velocity and mass ratio during vertical sphere water entry. *Dev. Appl. Oceanic Eng.*, **3**, 55–62.

Sen Gupta, A., N. Jourdain, J. Brown, and D. Monselesan, 2013: Climate drift in the CMIP5 models. *J. Climate*, **26**, 8597–8615, doi:10.1175/JCLI-D-12-00521.1.

Shepard, T., J. P. Abraham, D. S. Schwalbach, S. Kane, D. Sigling, and T. Harrington, 2014: Velocity and density effect on impact force during water entry of spheres. *J. Geophys. Remote Sens.*, **3**, 129, doi:10.4172/2169-0049.1000129.

Smith, D. M., and J. M. Murphy, 2007: An objective ocean temperature and salinity analysis using covariances from a global climate model. *J. Geophys. Res.*, **112**, C02022, doi:10.1029/2005JC003172.

——, and Coauthors, 2015: Earth's energy imbalance since 1960 in observations and CMIP5 models. *Geophys. Res. Lett.*, **42**, 1205–1213, doi:10.1002/2014GL062669.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Trenberth, K. E., 2015: Has there been a hiatus? *Science*, **349**, 691–692, doi:10.1126/science.aac9225.

——, and J. T. Fasullo, 2012: Tracking Earth's energy: From El Niño to global warming. *Surv. Geophys.*, **33**, 413–426, doi:10.1007/s10712-011-9150-2.

——, and ——, 2013: An apparent hiatus in global warming? *Earth's Future*, **1**, 19–32, doi:10.1002/2013EF000165.

——, ——, and M. Balmaseda, 2014a: Earth's energy imbalance. *J. Climate*, **27**, 3129–3144, doi:10.1175/JCLI-D-13-00294.1.

——, ——, G. Branstator, and A. S. Phillips, 2014b: Seasonal aspects of the recent pause in surface warming. *Nat. Climate Change*, **4**, 911–916, doi:10.1038/nclimate2341.

von Schuckmann, K., J.-B. Sallée, D. Chambers, P. Y. Le Traon, C. Cabanes, F. Gaillard, S. Speich, and M. Hamon, 2014: Monitoring ocean heat content from the current generation of global ocean observing systems. *Ocean Sci.*, **10**, 547–557, doi:10.5194/os-10-547-2014.

——, and Coauthors, 2016: An imperative to monitor Earth's energy imbalance. *Nat. Climate Change*, **6**, 138–144, doi:10.1038/nclimate2876.

Willis, J. K., D. Roemmich, and B. Cornuelle, 2004: Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *J. Geophys. Res.*, **109**, C12036, doi:10.1029/2003JC002260.

Wu, L., and Coauthors, 2012: Enhanced warming over the global subtropical western boundary currents. *Nat. Climate Change*, **2**, 161–166, doi:10.1038/nclimate1353.

Xue, Y., B. Huang, Z. Hu, A. Kumar, C. Wen, D. Behringer, and S. Nadiga, 2011: An assessment of oceanic variability in the NCEP climate forecast system reanalysis. *Climate Dyn.*, **37**, 2511–2539, doi:10.1007/s00382-010-0954-4.

——, and Coauthors, 2012: A comparative analysis of upper ocean heat content variability from an ensemble of operational ocean reanalyses. *J. Climate*, **25**, 6905–6929, doi:10.1175/JCLI-D-11-00542.1.

Zang, X. Y., and C. Wunsch, 2001: Spectral description of low-frequency oceanic variability. *J. Phys. Oceanogr.*, **31**, 3073–3095, doi:10.1175/1520-0485(2001)031<3073:SDOLFO>2.0.CO;2.

Zhang, Y., J. M. Wallance, and D. S. Battisti, 1997: ENSO-like interdecadal variability. *J. Climate*, **10**, 1004–1020, doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.