

Article

# A Simple Bias Correction Scheme in Ocean Data Assimilation

Changxiang Yan \* and Jiang Zhu

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

\* Correspondence: ycxlasg@mail.iap.ac.cn

**Abstract:** The mode bias is present and time-dependent due to imperfect configurations. Data assimilation is the process by which observations are used to correct the model forecast, and is affected by the bias. How to reduce the bias is an important issue. This paper investigates the roles of a simple bias correction scheme in ocean data assimilation. In this scheme, the misfits between modeled and monthly temperature and salinity with interannual variability from the Met Office Hadley Centre subsurface temperature and salinity data set (EN4.2.2) are used for the innovations in assimilation via the Ensemble Optimal Interpolation method. Two assimilation experiments are implemented to evaluate the impacts of bias correction. The first experiment is a data assimilation system without bias correction. In the second experiment, the bias correction is applied in assimilation. For comparison, the nature run with no assimilation and no bias correction is also conducted. When the bias correction is not applied, the assimilation alone leads to a rising trend in the heat and salt content that is not found in the observations. It is a spurious temporal variability due to the effect of the bias on the data assimilation. Meanwhile, the assimilation experiment without bias correction also produces significant negative impacts on the subsurface salinity. The experiment with bias correction performs best with notable improvements over the results of the other two experiments.

**Keywords:** bias correction; ensemble optimal interpolation; ocean heat and salt content; data assimilation



**Citation:** Yan, C.; Zhu, J. A Simple Bias Correction Scheme in Ocean Data Assimilation. *J. Mar. Sci. Eng.* **2023**, *11*, 205. <https://doi.org/10.3390/jmse11010205>

Academic Editors: Shaoqing Zhang, Yuxin Zhao, Hao Zuo and Junyu Dong

Received: 17 November 2022

Revised: 5 January 2023

Accepted: 9 January 2023

Published: 12 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ocean data assimilation combines observations with information from the numerical model to generate the accurate and comprehensive reanalyses of ocean states [1,2]. It is also used for the initialization of ocean forecasts [3,4]. The advantage of data assimilation in reducing the uncertainty and in improving forecasts has been presented [5–9]. However, state-of-the-art ocean general circulation models still suffer from model biases due to incorrect physical parameterizations and surface forcing, poor resolution, numerical dispersion, faulty boundary and initial conditions, as well as various other imperfections. Moreover, it is very difficult to determine the mechanisms underlying bias generation. All ocean data assimilation systems are affected by the bias. The problem of bias is an important issue in the reanalysis [10–14].

Considerable efforts have been made to remove model biases, and data assimilation is a common practice to reduce biases [13–16]. Dee and Da Silva [15] presented a bias estimation and correction algorithm in sequential data assimilation. It was successfully applied to the assimilation of humidity observations in the Goddard Earth Observing System [17]. Most studies augmented a model state vector by a bias vector to estimate and correct biases using data assimilation methods [16–20]. In some algorithms [18,19], the bias correction is not applied directly to the observed variable, but applied as a correction to the non-observed variable. The direct bias correction to the observed variable induced a spurious adjustment to other variables [19]. Additionally, the bias correction may be achieved by applying the cost function to fluctuations [21]. Moreover, the bias and the state vector may use different control variables [14,18,19]. In ocean data assimilation, temperature, salinity, current, and sea surface height are called control variables of the state vector. The variables which are bias corrected are called control variables of the

bias. The control variables of the bias may be same as those of the state vector, and also consist of salinity alone, or temperature alone, or different combinations among variables. Additionally, the control variables of the bias may also be other different variables such as pressure. Balmaseda et al. [19] proposed a multi-scale bias correction scheme, and it was successfully used for the ECMWF operational ensemble reanalysis–analysis system [14]. The above bias correction algorithms include a model depicting the time evolution of the bias. The bias model consisting of some bias parameters serves to predict the bias for time  $t_k$  based on a previous bias estimate valid a for time  $t_{k-1}$ . Then, the predicted bias is used to update bias estimate for time  $t_k$  using the observations via the method analogous to data assimilation. Its implementation requires specification of the error covariances of the bias estimate, background and observations, and an a priori bias estimation for time  $t_0$ . Finally, the bias update is corrected to the analysis field. However, there exists some uncertainty about an a priori bias estimation in the bias model. Meanwhile, the correct choice of the bias variables is important to the assimilation performance. Since the control variables of the bias may be different from those of the state vector, more choices of bias variables are available for a multivariate bias correction, and different choices may lead to totally different results. For example, the bias correction in temperature increased the velocity bias, while the bias correction in pressure reduced the bias in the velocity field [19]. It also implies that a bias needs to be attributed to a particular source. The wrong attribution will still force the assimilation to be consistent with a bias source, and lead to a wrong adjustment [19]. Additionally, the relationship among bias variables remains unclear. The resultant error covariance matrix for the bias vector is also uncertain. Moreover, in areas where few observations exist, it is also difficult for the above bias correction methods to work well. In this paper, we propose a simple bias correction scheme in ocean data assimilation using the Ensemble Optimal Interpolation (EnOI) method, and examine its impacts on the temperature and salinity. This paper is organized as follows: The model, bias, data and bias correction scheme are described in Section 2. The design and evaluation of the experiments are presented in Section 3. The discussion and conclusions are given in Section 4.

## 2. Model, Bias, Data and Bias Correction Scheme

### 2.1. Ocean Model

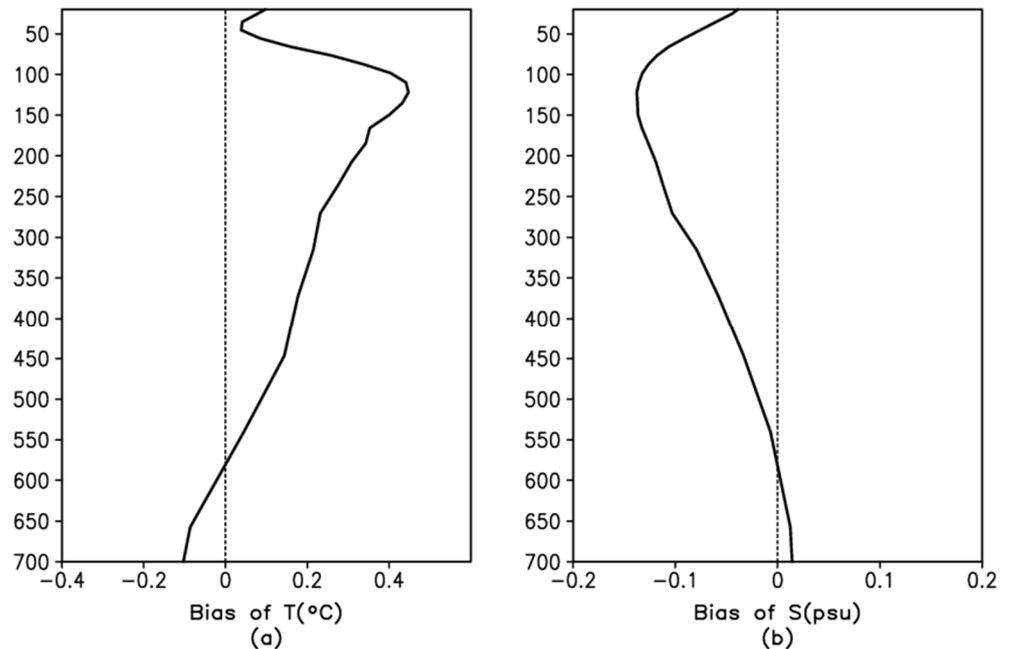
The model configuration used here is the hybrid coordinate ocean model (HYCOM) upgraded by the Nansen Environmental and Remote Sensing Center in Norway [22–24]. It uses three kinds of vertical coordinates according to the depth. The isopycnic vertical coordinates are adopted in the open, stratified ocean, and smoothly transit to z coordinates in the weakly stratified upper-ocean mixed layer, to terrain-following sigma coordinates in shallow water regions, and back to level coordinates in very shallow water. The hybrid coordinates are helpful to represent ocean thermodynamic processes and ocean flow better.

The model domain ranges from  $30^\circ$  E to  $70^\circ$  W, and from  $51^\circ$  S– $61^\circ$  N. The model horizontal resolution is approximately  $3/4^\circ \times 3/4^\circ$ . There are 28 hybrid layers, and the top five layers remain at the z-level coordinate with the minimum thickness of 2 m. The vertical mixing scheme is the K-profile parameterization [25]. The model temperature and salinity fields are relaxed toward the version 3.0 of the Polar Science Center Hydrographic Climatology (PHC) [26] with an e-folding time of 30 days at the lateral boundary. The surface forcing fields include the 6 h air temperature, dew temperature, winds, mean sea level pressure, total cloud cover, and precipitation from the atmospheric reanalysis ERA-interim [27].

### 2.2. Model Bias

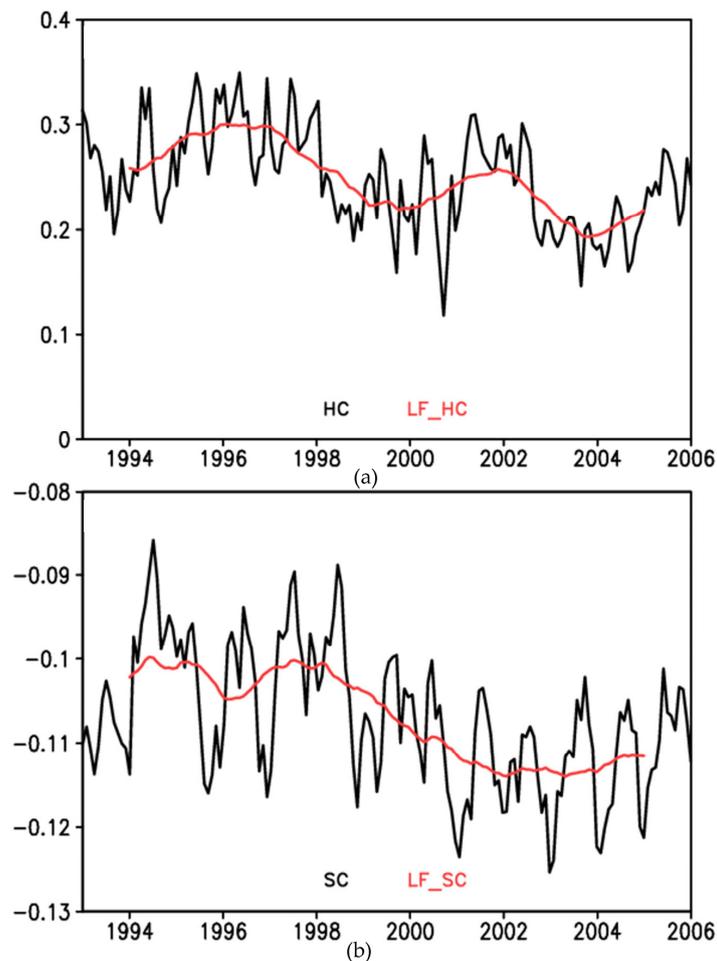
The nature run without assimilation and bias correction is carried out for the period of 1993–2005. The monthly temperature and salinity objective analyses from the Met Office Hadley Centre subsurface observations dataset EN4.2.2 using bias corrections of c14 [28–30] are used as independent observations to validate the experiment results. Figure 1 shows

vertical profiles of the 1993–2005 mean temperature and salinity difference between the simulations from the nature run and the observations, averaged over the model domain. The errors in the mean state provide evidence for the presence of bias. It is noticeable that the simulated temperature and salinity are biased with respect to the observations due to imperfect model configurations such as parameterization schemes, boundary conditions, forcing fields and coarse resolutions. The temperature is too warm with respect to the observations, while the salinity is too fresh. The warm bias reaches a maximum of 0.4 degrees in the thermocline. The salinity is fresher than the observations with a minimum bias of  $-0.14$  psu at about 120 m depth.



**Figure 1.** Vertical profiles of the 1993–2005 mean differences between the simulations and the observations, averaged over the model domain for (a) temperature, (b) salinity.

Figure 2 shows the time evolution of the errors of domain averaged temperature and salinity in the upper 300 m with respect to the observations. The 24-month running mean of the error, representative of the low frequency component, is shown in red. This component of the error is equivalent to the bias, and represents the part of the error that is correlated in time. The high frequency part is shown in black. Note that the magnitude of the low frequency part of the error is not small compared with the high frequency part. Furthermore, the low frequency component presents interannual variations. This result indicates that the bias is not constant in time. Particularly noticeable is the decreased trend after 1998 in the salinity bias. Meanwhile, the temperature bias also demonstrates a slight decreased trend in 1993–2005. The changes in the bias could be due to the errors in the thermocline. The large temperature gradient in the thermocline is difficult to simulate with a model that usually produces a flatter thermocline, especially during the cold phase of ENSO (El Niño-Southern Oscillation). Additionally, these may be associated with the surface heat and fresh water fluxes [31,32]. A better understanding of the changes needs more work, which is beyond the scope of this paper. It is notable that Figure 2 highlights the time-dependence of the bias, and it also suggests that data assimilation taking into account a bias correction is necessary.



**Figure 2.** Time series of the low and high frequency components of the errors for (a) temperature, and (b) salinity, averaged over the upper 300 m in the model domain. The 24-month running mean is shown in red and the high-frequency error is shown in black.

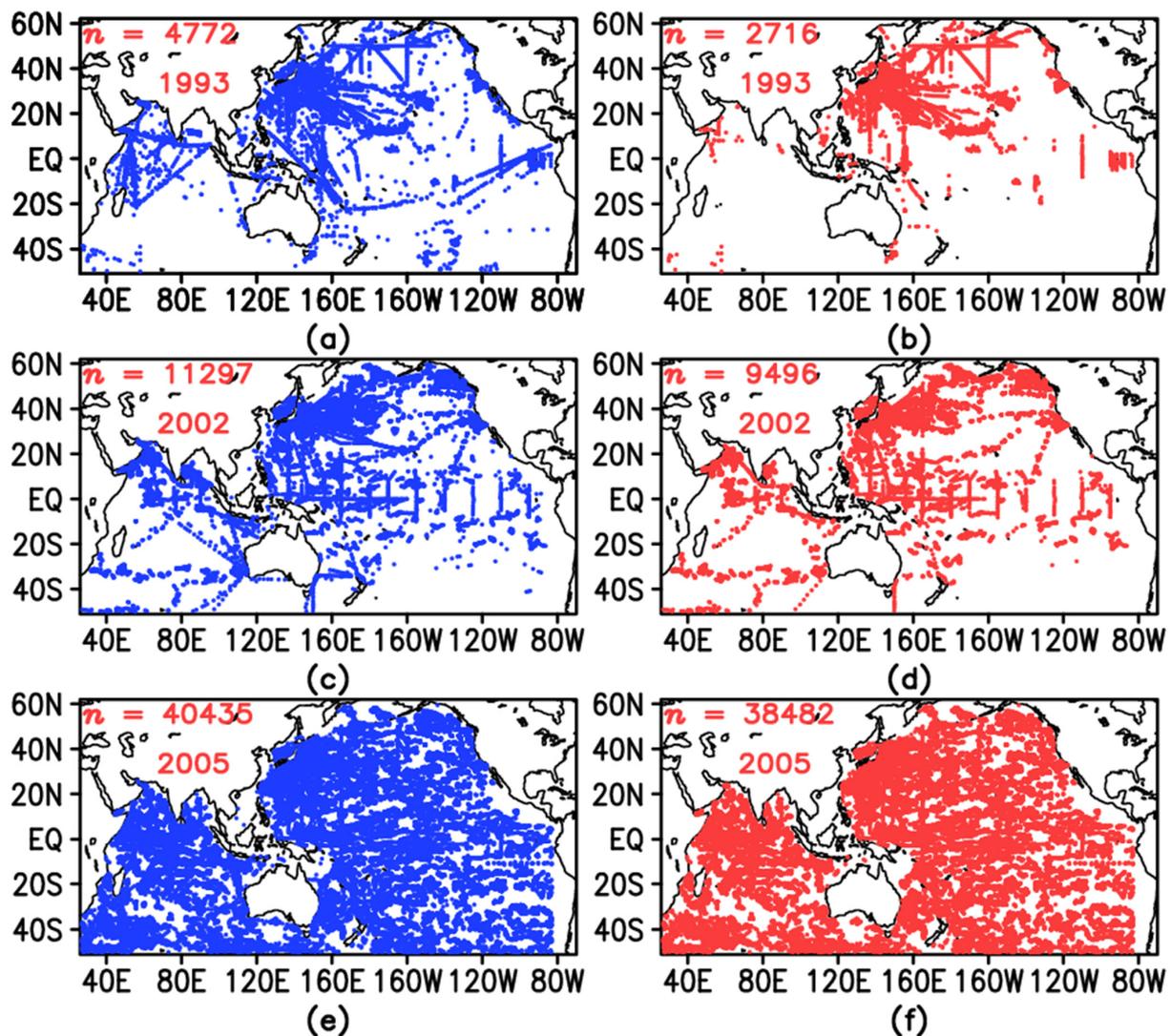
### 2.3. Data

The Sea Level Anomaly (SLA) data that are assimilated come from the global gridded product with a horizontal resolution of  $1/4^\circ \times 1/4^\circ$  and a temporal resolution of 1 day delivered by the Sea Level TAC (Thematic Assembly Centre) of the Copernicus Marine Environment Monitoring Service (CMEMS) project. It is generated by the Data Unification and Altimeter Combination System (DUACS) merging data from all altimeter missions (e.g., Sentinel-3A, Jason-3, HY-2A, Saral/AltiKa, Cryosat-2, OSTM/Jason-2, Jason-1, Topex/Poseidon, Envisat, GFO, ERS-1/2) [33]. The remotely sensed satellite sea surface temperature (SST) data that are assimilated come from the product of Reynolds et al. [34,35]. It was produced by combining the SST data from the Advanced Very High Resolution Radiometer (AVHRR) and Advanced Microwave Scanning Radiometer (AMSR) satellites with in situ data from ships and buoys using the optimum interpolation (OI) method at a spatial resolution of  $1/4^\circ \times 1/4^\circ$  and a temporal resolution of 1 day with global coverage [34].

The in situ temperature (T) and salinity (S) profiles assimilated come from the dataset EN4.2.2 [28]. The EN.4.2.2 data are obtained from <https://www.metoffice.gov.uk/hadobs/en4/> URL (accessed on 6 July 2022), and are © British Crown Copyright, Met Office, provided under a Non-Commercial Government Licence <http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/version/2/> URL (accessed on 6 July 2022). It includes four sources: Argo [36], the Arctic Synoptic Basin-wide Oceanography (ASBO) project [37,38], the Global Temperature and Salinity Profile Programme (GTSP) [39], and the World Ocean Database 2018 (WOD18) [40]. A new time-, temperature-, and probe-

type-dependent correction scheme, in which the pure thermal bias was assumed to be determined by both time and water temperature and the depth error was affected by calendar year and 0–100-m-averaged water temperature, was proposed to correct the historical eXpendable BathyThermograph (XBT) data [29]. A country-, depth-, and time-dependent correction scheme, in which the depth-dependent depth bias and the depth-independent pure thermal bias were considered, was implemented to correct Mechanical BathyThermograph (MBT) data [30]. This dataset made use of the results of quality control procedures that used altimeter data developed by Guinehut et al. [41].

The spatial and temporal distribution of in situ observations shows that the temperature and salinity profiles are sparse (Figure 3), confined to the observation means and observation cost. Moreover, they are very irregular. Before the Argo era, the spatial distribution of the observations was extremely inhomogeneous (Figure 3a,b). The observations are concentrated in the Northern Indian Ocean and the Northwestern Pacific Ocean, while there are few observations in the southern oceans (Figure 3a,b). The implementation of the Argo plan extends greatly the coverage of the observations except for the coastal seas (Figure 3c,d). Compared with the temperature observations, the salinity observations are scarcer, especially before the Argo project. The distribution highlights the changes in the observing system.



**Figure 3.** The distribution of the in situ temperature (a,c,e) and salinity (b,d,f) observations in different years. (a,b) 1993, (c,d) 2002, (e,f) 2005. The  $n$  in each figure denotes the number of the profile observations.

### 2.4. The Bias Correction Scheme

The satellite data provide nearly global coverage, while in situ temperature and salinity observations appear irregular and inhomogeneous in time and space. The ocean data assimilation without bias correction is not capable of working well due to the presence of the bias, particularly in regions where few observations exist. In this paper, a simple bias correction scheme is designed primarily for temperature and salinity by means of data assimilation.

The ocean data assimilation scheme used in this paper is an ensemble optimal interpolation (EnOI) method [42]. The analysis fields are given by solving the following equations:

$$\psi^a = \psi^b + \mathbf{W}(y - \mathbf{H}\psi^b) \tag{1}$$

$$\mathbf{W} = \frac{\alpha(\mathbf{C} \circ \mathbf{P})\mathbf{H}^T}{(\alpha\mathbf{H}(\mathbf{C} \circ \mathbf{P})\mathbf{H}^T + \mathbf{R})} \tag{2}$$

where  $\psi = [t, s, d, u, v, ub, vb, pb]^T$  represents the model state vector including temperature, salinity, layer thickness, baroclinic and barotropic current fields, and barotropic pressure. The superscripts a, b and T denote analysis, background, and matrix transpose, respectively.  $\mathbf{H}$  is the linear observation operator that interpolates from the model space to the observation space.  $\mathbf{P}$  is the background error covariance matrix.  $\mathbf{C}$  is a correlation function used to localize the background error covariance. Each element of  $\mathbf{C}$  is computed by a 5th-order piecewise rational function [43] where the length scale is taken as 400 km. The circle between  $\mathbf{C}$  and  $\mathbf{P}$  denotes a Schur product.  $\mathbf{R}$  is the observation error covariance matrix. Since the observation errors are usually assumed to be uncorrelated,  $\mathbf{R}$  is a diagonal matrix.  $\alpha$  is a scalar that decides the weights on the ensemble versus observations. Here it is taken as 0.4 [44].

The background error covariance matrix  $\mathbf{P}$  is estimated by

$$\mathbf{P} = \left( \frac{1}{n-1} \right) \mathbf{A}\mathbf{A}^T \tag{3}$$

where  $\mathbf{A} = (\psi_1, \psi_2, \dots, \psi_n)$  is an ensemble consisting of model states from the long-time model integration and  $n$  is the ensemble size ( $n = 120$  [45] here). To reflect the structure of the background error covariances well, different ensembles are employed for different seasons. In each season, the ensemble with the size of 120 is randomly sampled every 9 days over the 16-year model nature run in this season.

The EnOI may be used to assimilate SLA, SST and in situ T/S profiles. A different technique based on the EnOI is used to assimilate T/S profiles due to the isopycnic coordinate in HYCOM [46]. The layer thickness calculated from temperature and salinity observations is assimilated to adjust the model layer thickness and current fields. Then, the temperature or salinity observations are used to adjust the model temperature or salinity followed by diagnosing the salinity or temperature from the equation of the seawater state.

As shown in Figures 1 and 2, the model bias is present and time-dependent. The designed bias correction scheme takes this into account, and is performed during the above data assimilation. The temperature and salinity misfits (analysis-minus-simulation) between the simulations and the same monthly analysis with interannual variations from EN4.2.2 are regarded as the innovations. Only a fraction of temperature and salinity profiles sampled at regular intervals of 4 gridpoints from the analysis are used to calculate the misfits, since excessive analysis data may filter some high-frequency information from the observations and also increase computational resources. The remaining profiles are regarded as independent observations. At each assimilation step, the EN4.2.2 analysis-minus-background residuals are calculated. And then, the residuals are regarded as the innovations to join in the assimilation.

### 3. Results

#### 3.1. Experiment Setup

To examine the impacts of the bias correction scheme on the temperature and salinity, we carried out two assimilation experiments, in which all observations including SST, SLA and T/S profiles were assimilated into HYCOM, for the period 1993–2005.

In the first experiment, all observations are assimilated, but the bias correction is not applied (hereafter called Without\_Correct). In the second experiment, the bias correction is applied in the assimilation with all observations (hereafter called With\_Correct). The nature run (hereafter called CNTL) without any assimilation is carried out to examine if the bias correction scheme is capable of reducing the bias. Table 1 details the design of the experiments.

Table 1. Experiment setup.

Experiment Name	Scheme	Assimilated Data	Assimilation Frequency	Period
Without_Correct	No bias correction	SST, SLA, T/S	7 days	1993–2005
With_Correct	Bias correction	SST, SLA, T/S	7 days	1993–2005
CNTL	No assimilation and no bias correction	None	N/A	1993–2005

#### 3.2. Impacts on Subsurface Temperature and Salinity

It is natural that the assimilation may improve the modeled sea surface state due to the assimilation of satellite data with global coverage. Whether or not the subsurface temperature and salinity is improved is investigated in this section. The independent temperature and salinity objective analysis data from EN4.2.2 are used to validate the impacts of the bias correction in assimilation on the subsurface temperature and salinity. EN4.2.2 ingested data from all types of ocean profiling instruments that provide temperature and salinity information, and covered the period 1900 to present. Moreover, EN4.2.2 data were subjected to a series of quality control checks and bias adjustments. The objective analysis was calculated from the quality-controlled data using the Analysis Correction scheme [47]. Therefore, EN4.2.2 is very suitable for validation of temperature and salinity in this paper.

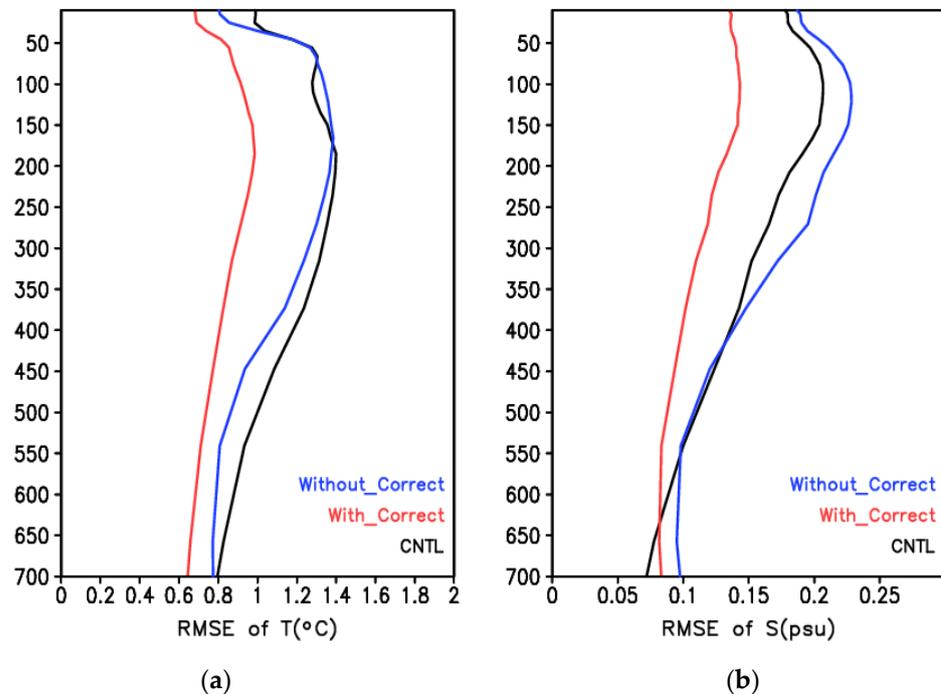
According to the spatial contribution of the temperature and salinity observations, the Southern Ocean with scarce data coverage shown in Figure 3 is used to examine the role of the bias correction. To further assess the performance of the bias correction, the Root Mean Square Error (RMSE) is used as an evaluation metric. The RMSE is defined as

$$RMSE(f, o) = \frac{1}{n} \sqrt{\sum_i^n (f - o)^2} \tag{4}$$

where  $f$  denotes the simulated results from different experiments,  $o$  denotes the observations, and  $n$  is the number of values in the data.

Figure 4 shows the vertical distribution of the RMSEs of temperature and salinity in the Southern Ocean over the period 1993–2005. For temperature, the experiment Without\_Correct, in which no bias correction is applied, produces an improvement over the nature run in the upper 50 m and below 200 m, and a slight decline in the thermocline. The SST assimilation contributes much to the sea surface improvement, while the T/S assimilation does much to the subsurface. The decline is likely associated with the limited temperature observations. The experiment With\_correct, in which bias correction is applied, exhibits a great improvement over the nature run in the entire water column. For salinity, Without\_Correct presents the worst performance with the largest RMSE among experiments. This result is related to sparse salinity data coverage. Most of observations are for temperature in the Southern Ocean before 2000, since more salinity observations were added to the Argo project. The long-term lack of salinity observations will induce a

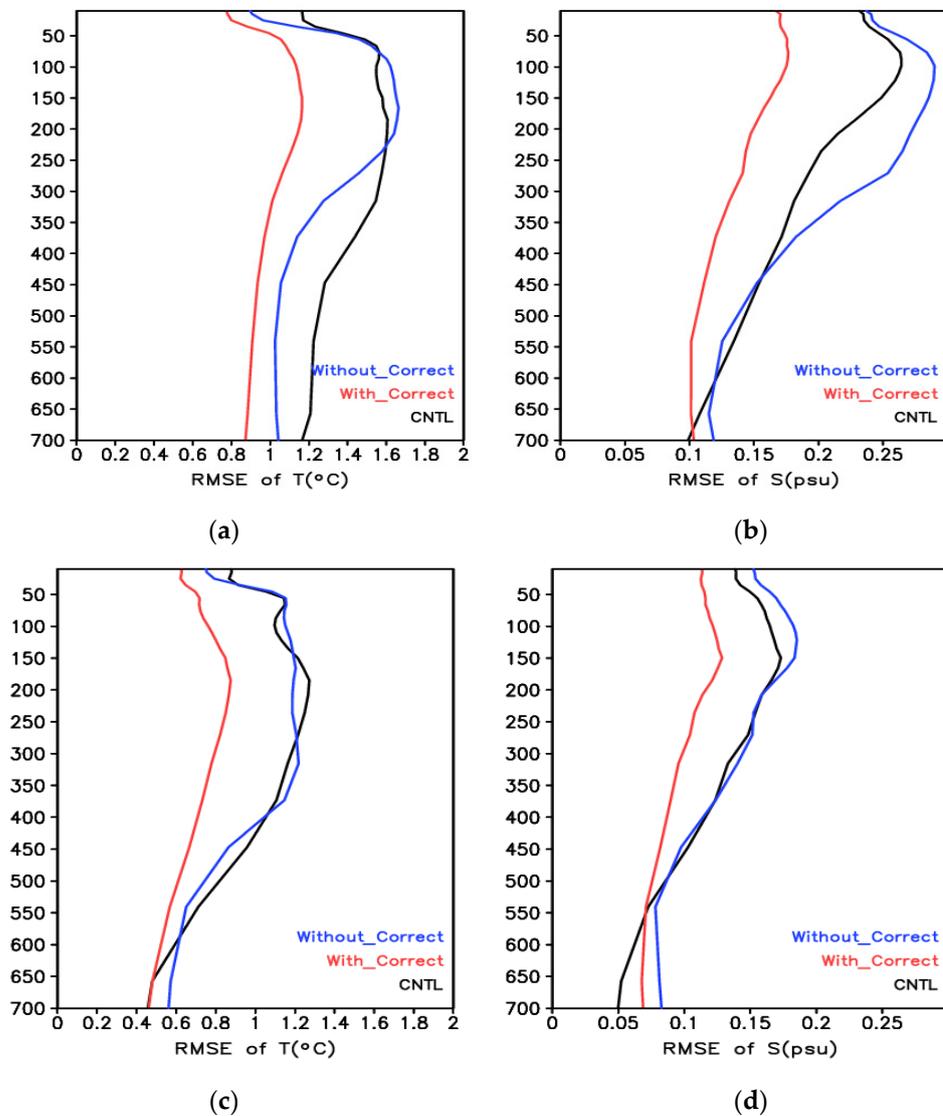
deviation. The experiment *With\_correct* observably demonstrates the best performance, as it indicates that the bias correction plays an important role for the ocean data assimilation.



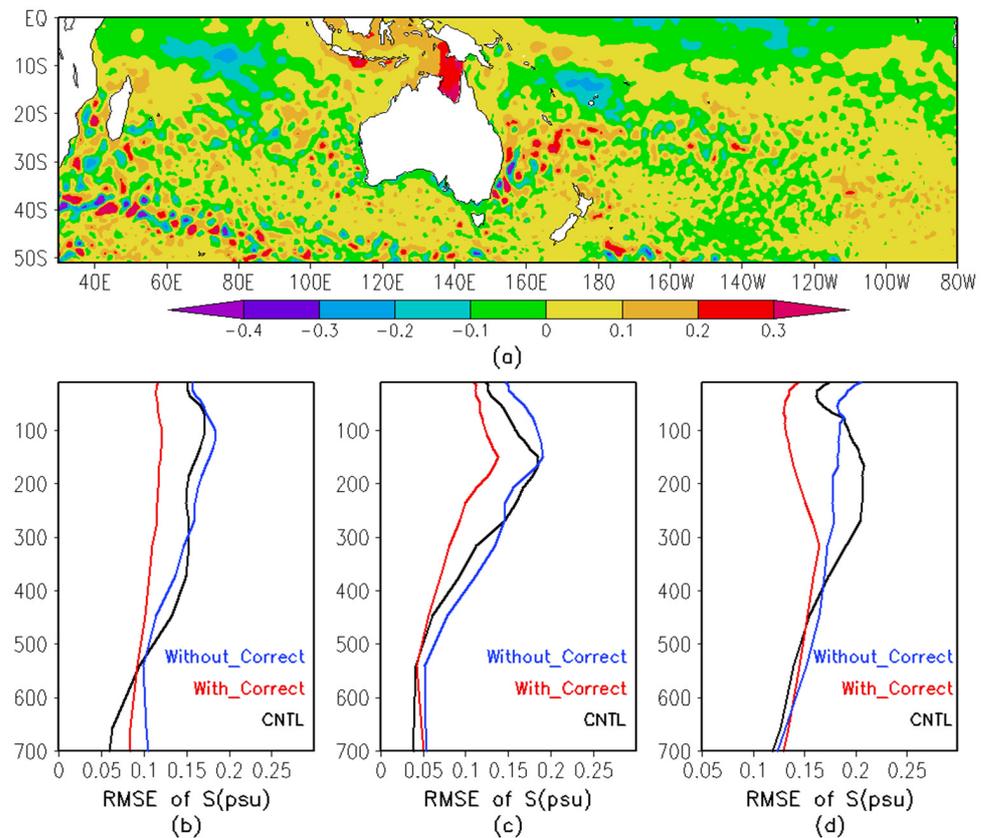
**Figure 4.** Vertical distribution of the RMSEs of (a) temperature and (b) salinity in the southern Indian and Pacific oceans over the period 1993–2005 for different experiments.

The vertical distribution of the RMSEs of temperature and salinity is different in different ocean basins (Figure 5). The experiment *With\_Correct* exhibits the best performance for temperature and salinity in the two ocean basins. For salinity, *Without\_Correct* performs worst in the two ocean basins with a salinity deterioration. Notably, *Without\_Correct* yields the largest RMSEs in the Southern Indian Ocean, especially in the upper 400 m, which contributes considerably to the total RMSEs in the Southern Ocean. For temperature, *Without\_Correct* produces an improvement in the sea surface, and a slight decline in the thermocline in the two ocean basins. Additionally, it presents a great improvement below the thermocline in the Southern Indian Ocean. This result is likely associated with the large modeled RMSEs. It is clear that the modeled RMSEs in the Southern Indian Ocean are larger than those in the Southern Pacific Ocean. The greater the model error is, the greater the adjustment induced by the assimilation is. In the southern Pacific Ocean, the salinity for both *Without\_Correct* and *With\_Correct* is degraded between 550 and 700 m, compared with CNTL. It is possibly associated with eddies. Abundant eddies are present (Figure 6a) in the southern Pacific Ocean. Most eddies are concentrated in the southern Pacific west of 150° W. Therefore, we divide the Southern Pacific Ocean into east and west parts by the 150° W line. The RMSEs of salinity from *Without\_Correct* and *With\_Correct* are larger than those of CNTL between 550 and 700 m in the two parts (Figure 6b,c). However, they are more remarkable to the west of 150° W, and contribute greatly to the total RMSEs in the Southern Pacific Ocean between 550 and 700 m. As shown by Figure 6a, plentiful eddies are also observed in the Southern Indian Ocean. We also validate the negative impact of the two assimilation experiments on the salinity in the subsurface (Figure 6d). However, the salinity between 550 and 700 m in the Southern Indian Ocean appears different from that in the Southern Pacific Ocean (Figure 5). It is possibly associated with larger adjustment to the background in the Indian Ocean (Figure 5). Those large correct adjustments compensate for the negative effects by eddies in the Southern Indian Ocean, while the small adjustments in the Southern Pacific Ocean do not offset the bad effects. The model has a coarse resolution

that is not capable of resolving meso-scale eddies, while the observations include eddies. The resultant ensemble used to estimate the background error covariance does not represent the correlation well in the regions with eddies. And thus, the assimilation leads to the slight decline of salinity in the 550 and 700 m.



**Figure 5.** Vertical distribution of the RMSEs of (a,c) temperature and (b,d) salinity for different ocean basins and different experiments over the period 1993–2005. (a,b) Southern Indian Ocean; (c,d) Southern Pacific Ocean.



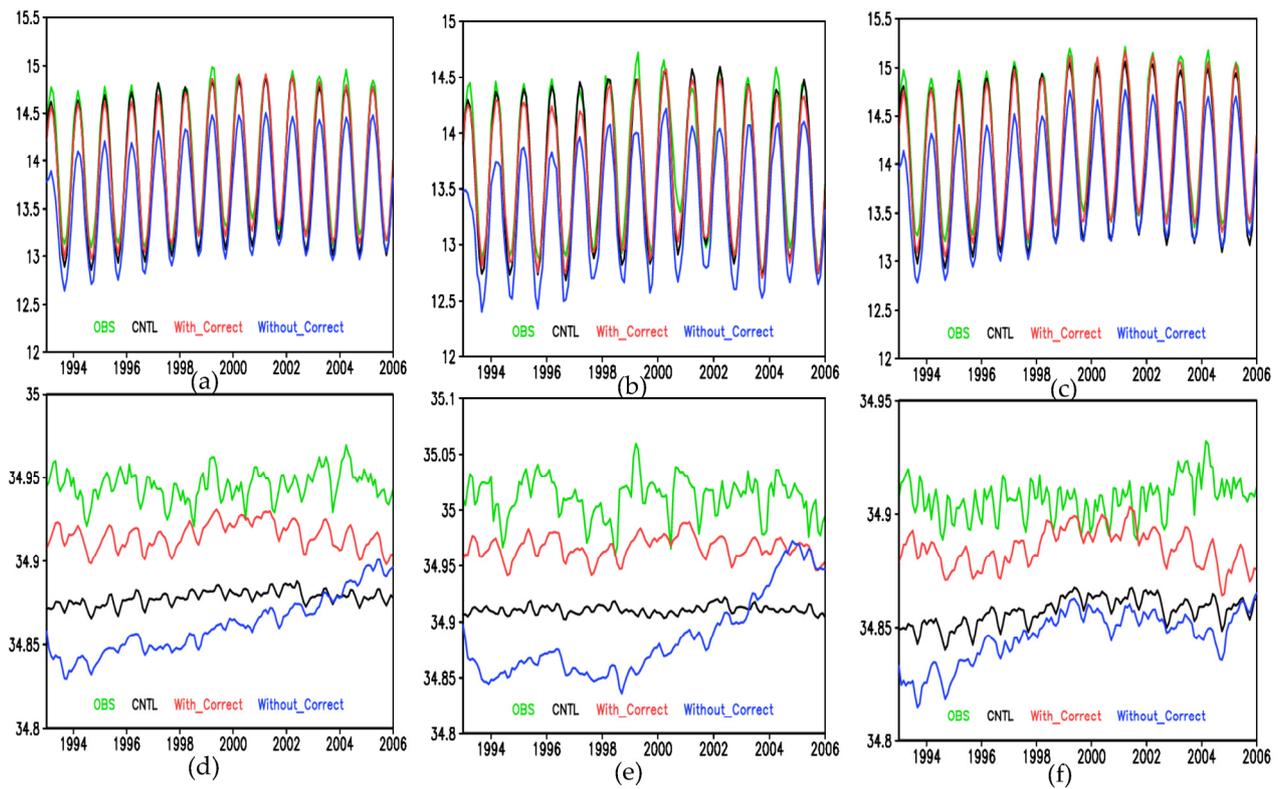
**Figure 6.** Observed SLA in Jan 1999 (a) and vertical distribution of the RMSEs of salinity (b–d) for different regions and different experiments over the period 1993–2005. (b) Southern Pacific Ocean west of 150° W, (c) Southern Pacific Ocean east of 150° W, (d) Southern Indian Ocean (30° E–40° E, 40° S–35° S).

### 3.3. Impacts on the Heat and Salt Content

The heat content represents the heat stored in the ocean, while the salt content represents the salt stored. They are important factors determining the subsurface temperature and salinity changes. The impacts of the bias correction on the heat content and salt content are investigated in this section.

Generally, the ocean heat content in the upper 300 m (HC300 hereafter) is defined as the temperature averaged over 0–300 m. Similarly, the ocean salt content in the upper 300 m (SC300 hereafter) is defined as the average salinity. HC300–700 represents the average temperature over 300–700 m, while SC300–700 represents the average salinity.

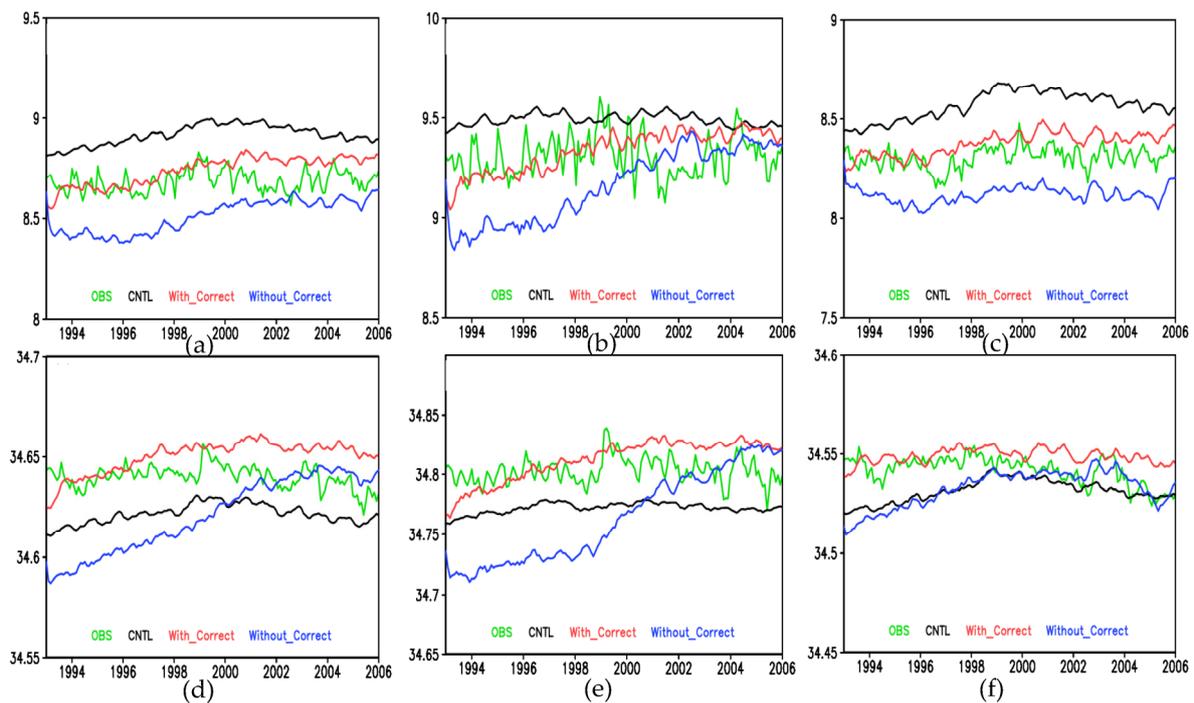
The HC300s averaged over the Southern Ocean (Figure 7) for each experiment demonstrate an annual cycle with the maximum in winter and the minimum in summer. It is associated with the solar radiation that is stronger in winter than in summer in the Southern Hemisphere. The HC300 from *With\_Correct* is relatively closer to the observation than that of *CNTL*. However, *Without\_Correct* shows the lowest HC300 among experiments, especially in the maximum value, which results in the largest difference from the observations. The SC300s averaged over the Southern Ocean from the three experiments are smaller than the observations. Overall, *With\_Correct* presents the closest SC300s to the observations among experiments, while *Without\_Correct* presents the farthest, particularly before 2002. Moreover, for SC300, both *With\_Correct* and *CNTL* exhibit flat trends consistent with those of the observations during the period 1993–2005, while *Without\_Correct* demonstrates a notably upward trend. This result implies that the lack of the long-time observations and the changes in the observing system may induce spurious time variability in the analysis when the bias correction is not applied.



**Figure 7.** Time series of HC300 (a–c) and SC300 (d–f) averaged in different oceans from different experiments and observations. (a,d) Southern Indian and Southern Pacific oceans, (b,e) Southern Indian Ocean, (c,f) Southern Pacific Ocean.

The HC300s and SC300s averaged in different ocean basins are different. In general, *With\_Correct* exhibits the best results among experiments for both temperature and salinity in the two ocean basins, while *Without\_Correct* presents the worst. The HC300 and SC300 differences between the experiment *Without\_Correct* and the observations are larger in the southern Indian Ocean than those in the southern Pacific Ocean. It may be associated with the scarcer observations in the southern Indian Ocean. For *Without\_Correct*, the SC300 differences from the observation are the largest among experiments in the 1990s in the southern Indian Ocean. It indicates that CNTL outperforms *Without\_Correct*. Note these differences drop so fast after the late 1990s that they are almost the same as those between the *With\_Correct* and the observation in 2005. This result contributes considerably to the total SC300 in the Southern Ocean. The observed SC300 is higher than that of the three experiments. Although the SC300 from *Without\_Correct* gradually increases and approaches the observations during 1993–2005, it is still smaller than that of CNTL in the southern Pacific Ocean, and it implies that the changes in the observing system have a great contribution to the results.

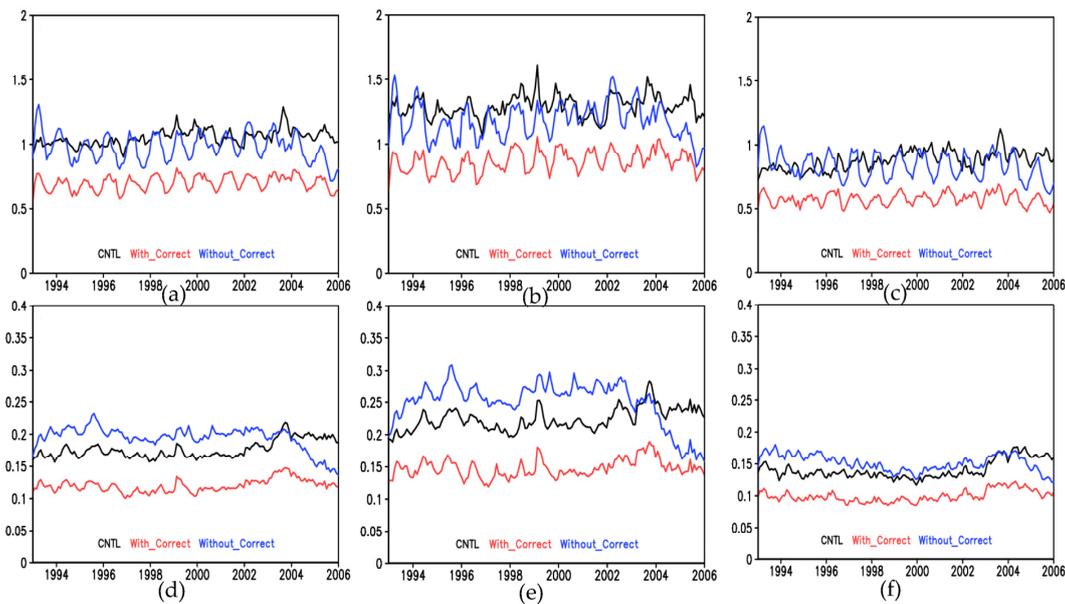
The HC300–700 (Figure 8a) from CNTL shows that the average temperature over 300–700 m is too high compared with the observation in the Southern Ocean, while the HC300–700 from *Without\_Correct* is too low. The data assimilation appears to overcorrect the model errors. The SC300–700 from CNTL is smaller than the observation (Figure 8d). The SC300–700 from *Without\_Correct* is smaller than that of CNTL in the 1990s. It indicates that the data assimilation makes negative contribution when the bias correction is switched off. Similar to SC300, the SC300–700 gradually approaches the observation with more available observations. *With\_Correct* shows the best performance for both HC300–700 and SC300–700. The results in different ocean basins are same as those in the Southern Ocean. For *Without\_Correct*, the offsets of HC300–700 and SC300–700 from the observation are larger in the Southern Indian Ocean than those in the Southern Pacific Ocean.



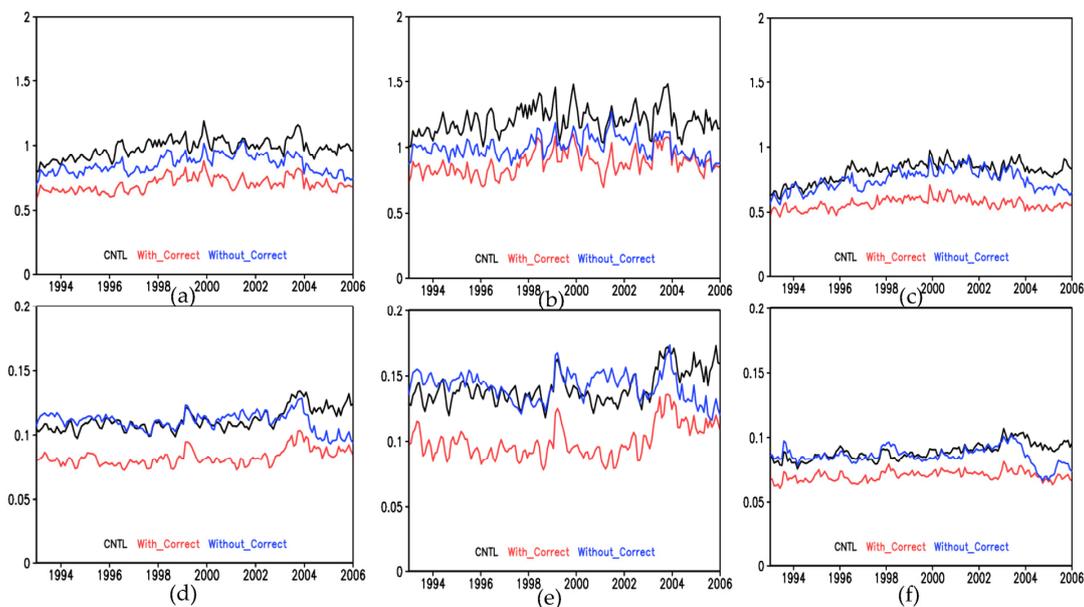
**Figure 8.** Time series of HC300–700 (a–c) and SC300–700 (d–f) averaged in different oceans from different experiments and observations. (a,d) Southern Indian and Southern Pacific oceans, (b,e) Southern Indian Ocean, (c,f) Southern Pacific Ocean.

Figure 9 shows the time evolution of the RMSEs of HC300 and SC300 relative to the independent analysis from EN4.2.2 over different ocean basins. For Southern Ocean, on average, the RMSE of HC300 from CNTL is basically 1 °C, while it is about 0.18 psu for SC300. It is clear that the bias correction has a prominent positive effect on both HC300 and SC300 (approximately reduction 30% in RMSE). The experiment without bias correction exhibits a slightly smaller RMSE of HC300 and a larger RMSE of SC300 compared with CNTL before 2004. When more Argo data are available, the RMSEs from Without\_Correct are reduced greatly. For different ocean basins, the time evolution of the RMSEs are similar to that in the Southern Ocean. As indicated by Figure 6, the three experiments exhibit the larger RMSEs in the Southern Indian Ocean than those in the Southern Pacific Ocean for HC300 and SC300. Meanwhile, compared with CNTL, the amplitude of the reduction or increasement in RMSEs from the two assimilation experiments is the largest in the Southern Indian Ocean, which makes great contribution to the total RMSEs in the Southern Ocean.

The RMSEs of both HC300–700 and SC300–700 from different experiments in different ocean basins (Figure 10) are smaller compared with those in the upper 300 m due to small changes in the subsurface. The larger RMSEs in the upper 300 m are induced primarily by the thermocline. For HC300–700 and SC300–700, With\_Correct produces the smallest RMSEs and performs best. It is noticeable that Without\_Correct produces a reduced RMSE of HC300–700, and a slightly increased RMSE of SC300–700 before 2003 compared with CNTL. After that, the RMSEs are reduced. Similar results are presented in different ocean basins. The RMSEs in 300–700 m are still larger in the Southern Indian Ocean than those in the Southern Pacific Ocean. Moreover, the magnitude of the reduction in RMSE is also larger compared with that in the Southern Pacific Ocean. As a result, the RMSEs in the Southern Indian Ocean contribute greatly to those in the Southern Ocean.



**Figure 9.** Time series of the RMSEs of HC300 (a–c) and SC300 (d–f) relative to EN4.2.2 in different ocean basins from different experiments. (a,d) Southern Indian and Southern Pacific oceans, (b,e) Southern Indian Ocean, (c,f) Southern Pacific Ocean.



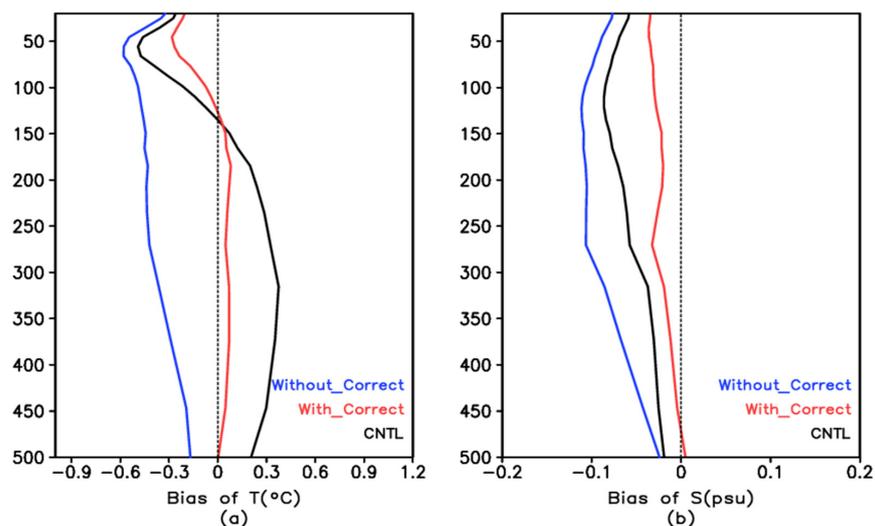
**Figure 10.** Time series of the RMSEs of HC300–700 (a–c) and SC300–700 (d–f) relative to EN4.2.2 in different ocean basins from different experiments. (a,d) Southern Indian and Southern Pacific oceans, (b,e) Southern Indian Ocean, (c,f) Southern Pacific Ocean.

#### 4. Discussion and Conclusions

This study addresses the impact of a simple bias correction scheme in ocean data assimilation on the temperature and salinity simulations based on the EnOI assimilation method. Three different experiments are carried out for the period 1993–2005, of which two are assimilation experiments (Without\_Correct, With\_Correct). Experiment Without\_Correct is a standard data assimilation analysis without bias correction. In With\_Correct, there is bias correction. A nature run is also used for comparison.

To better understand the roles of bias correction in assimilation, we choose the southern Indian and southern Pacific oceans with sparse observation coverage as the target regions. The bias correction in data assimilation results in a distinct improvement in temperature

and salinity. However, there is a negative impact on the temperature in the thermocline and on the salinity in the upper 400 m when the bias correction is not applied in data assimilation. The ocean heat and salt content in the upper 300 m is improved significantly by the bias correction in assimilation. A rising trend in the average SC300 and a much larger RMSE than that of CNTL are observed over time when bias correction is switched off in assimilation, particularly in the Southern Indian Ocean. It is a side effect of the standard data assimilation affected by model bias. Data assimilation systems influenced by bias are vulnerable to changes in the observing system. The ocean heat and salt content from 300 to 700 m reveals the changes in the subsurface temperature and salinity induced by the bias correction. The smallest RMSEs are observed when the bias correction is used in assimilation. For the heat content from 300 to 700 m, smaller RMSEs than those of the nature run without assimilation are found when no bias correction is applied in assimilation. The salt content from 300 to 700 m appears slightly worse in comparison with that of CNTL before more Argo data are available, and then has a sudden reduction in RMSE after that. This case is also illustrated by the RMSEs in the upper 300 m. The deterioration and sudden jump in RMSE of the salt content is a negative impact of the standard data assimilation without bias correction. The modeled thermocline is usually very diffuse, which implies the presence of large errors. The data assimilation tends to correct the temperature field by adding a negative increment. The vertical stability of the water column is disrupted due to no corresponding balance correction to the salinity field. As a result, the assimilation may lead to spurious adjustment to salinity. Figure 11 shows the biases of temperature and salinity over the period 1993–1999 in the Southern Ocean. For temperature, the nature run CNTL has a cold bias in the upper 150 m, and a warm bias below. After the bias correction is applied, both cold and warm biases are greatly reduced. When no bias correction is employed, the warm bias below 150 m is over-reduced so that it becomes a cold bias. For salinity, CNTL has a negative bias. Compared with CNTL, the bias correction still reduces the bias, while the bias of Without\_Correct is increased, and it indicates an incorrect adjustment. The results of this study suggest that the bias correction in ocean data assimilation is important, especially in regions with sparse observation coverage. The bias correction scheme used in this study is capable of treating time-dependent bias, and is relatively simple with no need to run the bias model. Moreover, it shows to be effective in improving the accuracy of the temperature and salinity simulations. Note this bias correction may reduce the bias, but not remove it all. Further work is needed to develop a satisfactory bias correction scheme for treatment of biases that are spatially variable, seasonal, diurnal and situation-dependent.



**Figure 11.** Vertical distribution of the biases of temperature (a) and salinity (b) over the period 1993–1999 for the southern Indian and southern Pacific oceans.

**Author Contributions:** Methodology, J.Z. and C.Y.; Data collection and treatment, C.Y.; Experiment setup, J.Z.; Result analysis, J.Z. and C.Y.; writing—original draft preparation, C.Y.; writing—review and editing, J.Z. and C.Y.; Discussion and conclusion, J.Z. and C.Y.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Strategic Priority Research Program of Chinese Academy of Sciences, grant number XDA19060102 and XDB42040106, the National Natural Science Foundation of China, grant number 41776041.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank CMEMS, Met Office Hadley Centre for the provided data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Artana, C.; Ferrari, R.; Bricaud, C.; Lellouche, J.M.; Garric, G.; Sennechal, N.; Lee, J.H.; Park, Y.H.; Provost, C. Twenty-five years of Mercator Ocean reanalysis GLORYS12 at Drake Passage: Velocity assessment and total volume transport. *Adv. Space Res.* **2021**, *68*, 447–466. [[CrossRef](#)]
2. Han, G.; Fu, H.; Zhang, X.; Li, W.; Wu, X.; Wang, X.; Zhang, L. A global ocean reanalysis product in the China Ocean reanalysis (CORA) project. *Adv. Atmos. Sci.* **2013**, *30*, 1621–1631. [[CrossRef](#)]
3. Pohlmann, H.; Jungclaus, J.; Marotzke, J.; Kohl, A.; Stammer, D. Improving predictability through the initialization of a coupled climate model with global oceanic reanalysis. *J. Clim.* **2009**, *22*, 3926–3938. [[CrossRef](#)]
4. Balmaseda, M.A.; Alves, O.J.; Arribas, A.; Awaji, T.; Behringer, D.W.; Ferry, N.; Fujii, Y.; Lee, T.; Rienecker, M.; Rosati, T.; et al. Ocean initialization for seasonal forecasts. *Oceanography* **2009**, *22*, 154–159. [[CrossRef](#)]
5. Hackert, E.C.; Kovach, R.M.; Busalacchi, A.J.; Ballabrera-Poy, J. Impact of Aquarius and SMAP satellite sea surface salinity observations on coupled ElNiño/Southern Oscillation forecasts. *J. Geophys. Res. Ocean.* **2019**, *124*, 4546–4556. [[CrossRef](#)]
6. Gao, C.; Zhang, R.; Wu, X.; Sun, J. Idealized experiments for optimizing model parameters using a 4D-Variational method in an intermediate coupled model of ENSO. *Adv. Atmos. Sci.* **2018**, *35*, 410–422. [[CrossRef](#)]
7. Zheng, F.; Zhu, J. Roles of initial ocean surface and subsurface states on successfully predicting 2006–2007 El Niño with an intermediate coupled model. *Ocean Sci.* **2015**, *11*, 187–194. [[CrossRef](#)]
8. Zhao, M.; Hendon, H.; Alves, O.; Yin, Y. Impact of improved assimilation of temperature and salinity for coupled model seasonal forecasts. *Clim. Dyn.* **2014**, *42*, 2565–2583. [[CrossRef](#)]
9. Lellouche, J.M.; Greiner, E.; Le Galloudec, O.; Garric, G.; Regnier, C.; Drevillon, M.; Benkiran, M.; Testut, C.E.; Bourdalle-Badie, R.; Gasparin, F.; et al. Recent updates on the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high resolution system. *Ocean Sci.* **2018**, *14*, 1093–1126. [[CrossRef](#)]
10. Eden, C.; Greatbatch, R.J.; Boning, C.W. Adiabatically correcting an eddy-permitting model using large-scale hydrographic data: Application to the Gulf Stream and the North Atlantic Current. *J. Phys. Oceanogr.* **2004**, *34*, 701–719. [[CrossRef](#)]
11. Chepurin, G.A.; Carton, J.; Dee, D.P. Forecast model bias correction in ocean data assimilation. *Mon. Wea. Rev.* **2005**, *133*, 1328–1342. [[CrossRef](#)]
12. Dee, D.P.; Uppala, S. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. Roy. Meteorol. Soc.* **2009**, *135*, 1830–1841. [[CrossRef](#)]
13. Dee, D.P. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 3323–3343. [[CrossRef](#)]
14. Zuo, H.; Balmaseda, M.A.; Tietsche, S.; Mogensen, K.; Mayer, M. The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Sci.* **2019**, *15*, 779–808. [[CrossRef](#)]
15. Dee, D.P.; Da Silva, A.M. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.* **1998**, *124*, 269–295. [[CrossRef](#)]
16. Balmaseda, M.A.; Mogensen, K.; Weaver, A.T. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.* **2013**, *139*, 1132–1161. [[CrossRef](#)]
17. Dee, D.P.; Todling, R. Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Mon. Wea. Rev.* **2000**, *128*, 3268–3282. [[CrossRef](#)]
18. Bell, M.J.; Martin, M.J.; Nichols, N.K. Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. R. Meteorol. Soc.* **2004**, *130*, 873–893. [[CrossRef](#)]
19. Balmaseda, M.A.; Dee, D.; Vidard, A.; Anderson, D.L.T. A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans. *Q. J. R. Meteorol. Soc.* **2007**, *133*, 167–179. [[CrossRef](#)]
20. Fujii, Y.; Nakaegawa, T.; Matsumoto, S.; Yasuda, T.; Yamanaka, G.; Kamachi, M. Coupled climate simulation by constraining ocean fields in a coupled model with ocean data. *J. Clim.* **2009**, *22*, 5541–5557. [[CrossRef](#)]

21. Shapiro, G.I.; Gonzalez-Ondina, J.M.; Salim, M.; Tu, J.; Asif, M. Crisis Ocean Modelling with a Relocatable Operational Forecasting System and Its Application to the Lakshadweep Sea (Indian Ocean). *J. Mar. Sci. Eng.* **2022**, *10*, 1579. [[CrossRef](#)]
22. Bleck, R. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean. Model.* **2002**, *4*, 55–88. [[CrossRef](#)]
23. Bertino, L.; Lisaeter, K.A.; Scient, S. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. *J. Oper. Oceanogr.* **2008**, *1*, 15–18. [[CrossRef](#)]
24. Sakov, P.; Counillon, F.; Bertino, L.; Lisaeter, K.A.; Oke, P.R.; Korablev, A. TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.* **2012**, *8*, 633–656. [[CrossRef](#)]
25. Large, W.G.; McWilliams, J.C.; Doney, S.C. Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* **1994**, *32*, 363–403. [[CrossRef](#)]
26. Steele, M.; Morley, R.; Ermold, W. PHC: A global ocean hydrography with a high-quality Arctic Ocean. *J. Clim.* **2001**, *14*, 2079–2087. [[CrossRef](#)]
27. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.* **2011**, *137*, 553–597. [[CrossRef](#)]
28. Good, S.A.; Martin, M.J.; Rayner, N.A. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Ocean* **2013**, *118*, 6704–6716. [[CrossRef](#)]
29. Cheng, L.; Zhu, J.; Cowley, R.; Boyer, T.; Wijffels, S. Time, Probe Type, and Temperature Variable Bias Corrections to Historical Expendable Bathythermograph Observations. *J. Atmos. Ocean. Technol.* **2014**, *31*, 1793–1825. [[CrossRef](#)]
30. Gouretski, V.; Cheng, L. Correction for Systematic Errors in the Global Dataset of Temperature Profiles from Mechanical Bathythermographs. *J. Atmos. Technol. Ocean. Res.* **2020**, *37*, 841–855. [[CrossRef](#)]
31. Zhang, R.H.; Wang, G.H.; Chen, D.; Busalacchi, A.J.; Hackert, E.C. Interannual biases induced by freshwater flux and coupled feedback in the tropical Pacific. *Mon. Weather Rev.* **2010**, *138*, 1715–1737. [[CrossRef](#)]
32. Hackert, E.; Ballabrera Poy, J.; Busalacchi, A.J.; Zhang, R.H.; Murtugudde, R. Impact of sea surface salinity assimilation on coupled forecasts in the tropical Pacific. *J. Geophys. Res.* **2011**, *116*, C05009. [[CrossRef](#)]
33. Pujol, M.I.; Faugere, Y.; Taburet, G.; Dupuy, S.; Pelloquin, C.; Ablain, M.; Picot, N. DUACS DT2014: The new multi-mission altimeter data set reprocessed over 20 years. *Ocean Sci.* **2016**, *12*, 1067–1090. [[CrossRef](#)]
34. Reynolds, R.W.; Smith, T.M.; Liu, C.Y.; Chelton, D.B.; Casey, K.S.; Schlax, M.G. Daily high-resolution blended analyses for sea surface temperature. *J. Clim.* **2007**, *20*, 5473–5496. [[CrossRef](#)]
35. Reynolds, R.W.; Banzon, V.F.; NOAA CDR Program. *NOAA Optimum Interpolation 1/4 Degree Daily Sea Surface Temperature (OISST) Analysis, Version 2*; NOAA National Centers for Environmental Information: Asheville, NC, USA, 2008.
36. Argo. *Argo Float Data and Metadata from Global Data Assembly Centre (Argo GDAC)*; SEANO: La Spezia, Italy, 2000.
37. Curry, R. *HydroBase 2—A Database of Hydrographic Profiles and Tools for Climatological Analysis*; Woods Hole Oceanographic Institution: Falmouth, MA, USA, 2001.
38. Swift, J. *Physical and Chemical Properties from Selected Expeditions in the Arctic Ocean*; Digital Media; Dichtl, R., Axford, Y., Haran, T., Eds.; National Snow and Ice Data Center: Boulder, CO, USA, 2002.
39. Sun, C.; Thresher, A.; Keeley, R.; Hall, N.; Hamilton, M.; Chinn, P.; Tran, A.; Goni, G.; Petit de la Villeon, L.; Carval, T.; et al. The data management system for the global temperature and salinity profile programme. In *Proceedings of the OceanObs.09: Sustained Ocean Observations and Information for Society*, Venice, Italy, 21–25 September 2009; Hall, J., Harrison, D.E., Stammer, D., Eds.; ESA Publication WPP-306: Paris, France, 2010; Volume 2. [[CrossRef](#)]
40. Boyer, T.P.; Antonov, J.I.; Baranova, O.K.; Coleman, C.; Garcia, H.E.; Grodsky, A.; Johnson, D.R.; Locarnini, R.A.; Mishonov, A.V.; O'Brien, T.D.; et al. *World Ocean Database 2013*; Levitus, S., Mishonov, A., Eds.; NOAA Atlas NESDIS 72: Washington, DC, USA, 2013; p. 209.
41. Guinehut, S.; Coatanoan, C.; Dhomp, A.-L.; Le Traon, P.-Y.; Larnicol, G. On the use of satellite altimeter data in Argo quality control. *J. Atmos. Ocean. Technol.* **2009**, *26*, 395–402. [[CrossRef](#)]
42. Evensen, G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean. Dyn.* **2003**, *53*, 343–367. [[CrossRef](#)]
43. Gaspari, G.; Cohn, S.E. Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteor. Soc.* **1999**, *125*, 723–757. [[CrossRef](#)]
44. Yan, C.; Zhu, J.; Tanajura, C.A.S. Impacts of mean dynamic topography on a regional ocean assimilation system. *Ocean Sci.* **2015**, *11*, 829–837. [[CrossRef](#)]
45. Zhu, J.; Zheng, F.; Li, X. A new localization implementation scheme for ensemble data assimilation of non-local observations. *Tellus A* **2011**, *63*, 244–255. [[CrossRef](#)]

46. Xie, J.; Zhu, J. Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. *Ocean Model.* **2010**, *33*, 283–298. [[CrossRef](#)]
47. Lorenc, A.C.; Bell, R.S.; Macpherson, B. The Meteorological Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.* **1991**, *117*, 59–89. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.