



Article

Parameter Estimation Based on a Local Ensemble Transform Kalman Filter Applied to El Niño–Southern Oscillation Ensemble Prediction

Yanqiu Gao ^{1,2} , Youmin Tang ^{3,4,*} , Xunshu Song ^{1,2} and Zheqi Shen ^{2,3}

¹ State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China; gaoyanqiu@sio.org.cn (Y.G.); songxs@sio.org.cn (X.S.)

² Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai 519082, China; zqshen@hhu.edu.cn

³ College of Oceanography, Hohai University, Nanjing 210098, China

⁴ Environmental Science and Engineering, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada

* Correspondence: ytang@unbc.ca

Abstract: Parameter estimation plays an important role in reducing model error and thus is of great significance to improve the simulation and prediction capabilities of the model. However, due to filtering divergence, parameter estimation by ensemble-based filters still faces great challenges. Previous studies have shown that a covariance inflation scheme could alleviate the filtering divergence problem by increasing the signal-to-noise ratio of the state-parameter covariance. In this study, we proposed a new inflation scheme based on a local ensemble transform Kalman filter (LETKF). With the new scheme, the Zebiak–Cane (Z-C) model parameters were estimated by assimilating the sea surface temperature anomaly (SSTA) data. The effectiveness of the parameter estimation and its influence on El Niño–Southern Oscillation (ENSO) prediction were evaluated in an observation system simulation experiments (OSSE) framework and real-world scenario, respectively. With the utilization of the OSSE framework, the results showed that the model parameters were successfully estimated. Parameter estimation reduced the model error when compared with only state estimation (onlySE); however, multiple parameter estimation (MPE) further improved the ENSO prediction skill by providing better initial conditions and parameter values than the single parameter estimation (SPE). Parameter estimation could thus alleviate the spring prediction barrier (SPB) phenomenon of ENSO to a certain extent. In real-world experiments, the optimized parameters significantly improved the ENSO forecasting skill, primarily in prediction of warm events. This study provides an effective parameter estimation strategy to improve climate models and further climate predictions in the real world.

Keywords: inflation scheme; parameter estimation; LETKF; Zebiak–Cane model; ENSO prediction skill



Citation: Gao, Y.; Tang, Y.; Song, X.; Shen, Z. Parameter Estimation Based on a Local Ensemble Transform Kalman Filter Applied to El Niño–Southern Oscillation Ensemble Prediction. *Remote Sens.* **2021**, *13*, 3923. <https://doi.org/10.3390/rs13193923>

Academic Editor: Sergey M. Varlamov

Received: 8 August 2021

Accepted: 27 September 2021

Published: 30 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Many studies have shown that numerical climate prediction models, even the state-of-the-art coupled models, still yield significant errors in predicting the true state of the atmosphere and ocean. Apart from the significant impact of the initial conditions [1–3], model errors greatly reduce the accuracy of climate prediction [4–7]. Generally, model errors stem from three aspects: misfit of the dynamical core, approximation scheme of the physical processes, and errors from the model parameters [8,9]. Due to the difficulty in improving the first two sources, many studies are devoted to reducing model parameter errors [10–12]. Although the parameters cannot be observed directly, their uncertainty can be constrained by observational information. The process of optimizing parameters by assimilating observed data is known as parameter estimation [13,14].

Observation of the ocean and atmosphere has been greatly enhanced in recent decades, such as on site observation and remote sensing observation, which provide a great deal of data for data assimilation. Additionally, the study of parameter estimation in atmospheric, oceanic, or coupled models is evolving rapidly in the development of assimilation methods [15–18]. Parameter estimation is mainly implemented by two methods: the variational method [10,18–23] and ensemble Kalman filter (EnKF) [11,12,21,24,25]. Deriving the linear tangent and adjoint model, the four-dimensional variational method is difficult, even impractical, when implemented in fully coupled circulation models [19,21]. Therefore, EnKF and its variants are widely used, from simple to fully coupled circulation models. Many studies have shown that parameter estimation based on EnKF and its variants can successfully restrain the uncertainty of parameterization schemes, reduce the model bias, and thus improve the predictability of climate models [12,25–28]. However, most of these studies were conducted in perfect scenarios where the truth parameters are known (i.e., assuming that model errors are caused by erroneous parameters). It is a great challenge to perform parameter estimation in the real world, where the true parameters are typically unknown.

Theoretically, any model parameter in conjunction with state variables could be estimated using data assimilation. In practical application, parameter estimation is usually performed by state vector augmentation techniques because there is no direct observation of the parameters [29,30]. Thus, the accuracy of the parameter estimation tends to depend on the quality of the state-parameter covariance [31]. Due to the characteristics of constants or slow changes in the parameters and the limitation of the ensemble samples, the relatively small ensemble spread will further decrease with the assimilation process when using EnKF to estimate parameters for real-world applications; thus, making it difficult for the observations to continue to update the parameter ensembles. This phenomenon is referred to as the filtering divergence problem [32–34]. Filtering divergence impacts the estimate of the state-parameter covariance and degrades the accuracy of the parameter estimation. To enhance the signal-to-noise ratio of the covariance is a challenge for parameter estimation, especially for coupled parameter estimation in the real world [35]. One effective solution to prevent filtering divergence is by the covariance inflation scheme, which increases the signal-to-noise ratio of the covariance to a certain extent to improve the accuracy of the parameter estimation [11,12,33,36–39].

In this study, a new covariance inflation scheme is proposed based on the local ensemble transform Kalman filter (LETKF). With the Zebiak–Cane (Z-C) model, parameter estimation and its impact on El Niño–Southern Oscillation (ENSO) prediction are investigated in an observation system simulation experiments (OSSE) framework and real-world experiments, respectively. The remainder of this paper is arranged as follows: Section 2 briefly introduces the Z-C model, LETKF-based parameter estimation algorithm, parameters to be estimated, and the experimental designs in the OSSE framework and real-world experiments, separately. Results of parameter estimation and ENSO prediction are reported in Section 3. Finally, discussions and conclusions are given in Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Zebiak–Cane Model

The Z-C model, which is a well-known ocean–atmosphere coupled model for the tropical Pacific Ocean, has been widely applied for ENSO study. The description of the model parallels that of Gao et al. (2020) [3], as follows in this section. The atmospheric dynamics follows the Gill model [40], with steady-state, linear shallow-water equations. If the sea surface temperature anomaly (SSTA) with ENSO characteristics, which refer to the positive or negative SSTA in the equatorial central to eastern Pacific, are given, the model can generate the sea surface wind anomaly needed to drive the tropical Pacific Ocean. The oceanic dynamics, which are described by a reduced-gravity model, can generate the SSTA in the tropical Pacific Ocean forced by the sea surface wind stress anomaly. The oceanic thermodynamics describe the sea surface temperature (SST) anomalies and heat flux change using a three-dimensional nonlinear equation. The atmospheric model spans

over the domain of 101.25° E–286.875° E and 29° S–29° N, with a grid spacing of 5.625° (longitude) × 2° (latitude). The oceanic dynamics spans over the domain of 125° E–281° E and 28.75° S–28.75° N, with a grid spacing of 2° (longitude) × 0.5° (latitude). The oceanic thermodynamics cover the domain of 129.375° E–275.625° E and 19° S–19° N, with the same grid as that of the atmospheric model. The model time step is 10 days. More details about the Z-C model can be found in Zebiak and Cane (1987) [41]. Its latest version LDEO5 is adopted in this study. The model domain is shown in Figure 1.

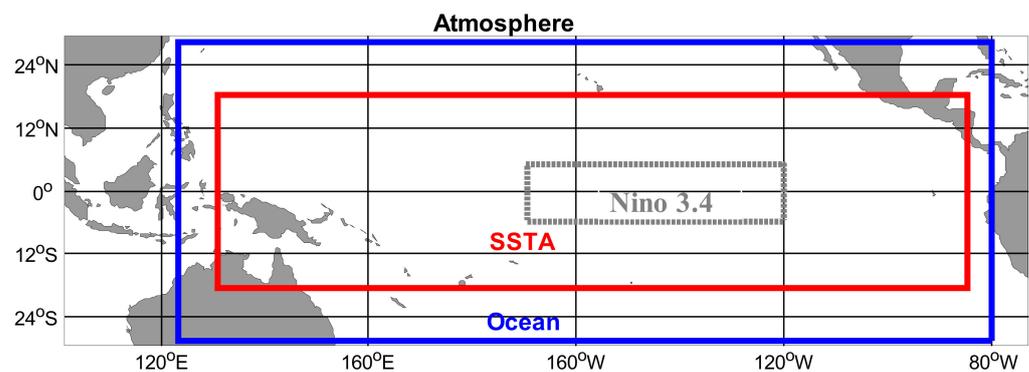


Figure 1. Integral domain of the LDEO5 model, including the atmospheric model contoured by solid black lines, oceanic model by blue lines, and SSTA model by red lines. The Niño3.4 region is also labeled by dotted grey lines.

2.2. LETKF-Based Parameter Estimation

As a variant of EnKF, LETKF calculates the inverse matrix on the ensemble space when calculating the analysis error covariance, which greatly saves computational expense [34]. This method has been preliminarily applied in the estimation of the model parameters, which reduces the analysis error and improves the simulation capability of the model [39,42]. Thus, we select LETKF to perform the state and parameter estimation in this study. The calculation steps of LETKF are listed as follows.

Step 1: Calculate the perturbation of the background in observation space using $\tilde{H} = H(x^b) - H(\bar{x}^b)$, where x^b and \bar{x}^b denote the forecast of the state and their ensemble means, respectively; H denotes the observation operator mapping the model state to observation space.

Step 2: Calculate the projection of the analysis error covariance matrix into the ensemble space using $\tilde{P}^a = [\tilde{H}^T R^{-1} \tilde{H} + (Ne - 1)I]^{-1}$, where R is the observation error covariance matrix, Ne represent the ensemble size, and I is a unit matrix.

Step 3: Calculate the Kalman gain matrix using $K = X^b \tilde{P}^a \tilde{H}^T R^{-1}$, where X^b denotes perturbations of the predictions from \bar{x}^b .

Step 4: The analysis x^a can be composed of two parts: Equation (1) analyzes the mean as \bar{x}^a , and Equation (2) analyzes the perturbation as X^a , which are briefly introduced here.

$$\bar{x}^a = \bar{x}^b + K(y^o - H(\bar{x}^b)), \quad (1)$$

$$X^a = X^b [(Ne - 1)\tilde{P}^a]^{1/2}, \quad (2)$$

where y^o represents observations. With a state vector augmentation technique, both \bar{x}^b and X^b include states and parameters, so that parameters can be estimated by Equations (1) and (2). It should be noted that \tilde{H} and \tilde{P}^a are only computed in Steps 1 and 2 in the state component. Appendix A provides details of the parameter estimation algorithm based on the state vector augmentation technique.

2.3. Covariance Inflation Scheme

In general, the covariance inflation scheme is implemented through inflating posterior perturbations of the parameters [43], which can be formulated as:

$$\tilde{\beta}_i = \bar{\beta} + \gamma(\beta_i - \bar{\beta}), \quad (3)$$

where $\tilde{\beta}_i$ and β_i denote inflated and posterior parameters, respectively. $\bar{\beta}$ represents the ensemble mean of β_i . γ is called the inflation factor. In previous studies, several inflation schemes were proposed according to different inflation factors, such as the fixed inflation (FI) scheme, conditional covariance inflation (CCI) scheme, and estimated parameter ensemble spread (EPES) scheme. In addition to the above schemes applied in parameter estimation, there are also some schemes applied in state estimation, such as the relaxation-to-prior perturbations (RTPP) scheme and relaxation-to-prior spread (RTPS) scheme. The inflation factors in these schemes are shown in Appendix B. However, most of these inflation schemes take a machinery strategy, namely, the inflation is continuously running during all assimilation steps, resulting in the parameters estimated to converge easily to the wrong values due to over-inflated model errors. In the CCI scheme, the posterior parameter ensemble is inflated only if its ensemble spread is smaller than the threshold.

An ideal inflation scheme should balance under-estimation and over-estimation of model error covariance; i.e., inflation is performed only when necessary. Thus, we propose a new covariance inflation scheme based on the CCI scheme. In the proposed scheme, the posterior ensemble perturbations are inflated at the beginning of parameter estimation, but unchanged when the parameter ensemble spread is smaller than the threshold. The inflation factor in Equation (3) can be written as

$$\gamma = \begin{cases} \frac{a}{\sigma_\beta}, & \sigma_\beta \geq b; \\ 1, & \sigma_\beta < b. \end{cases} \quad (4)$$

where σ_β is the posterior ensemble standard deviation, a is a fixed factor, and b is the threshold. Considering that the posterior ensemble standard deviation is constantly changing with the assimilation in nature, the factors of a and b in Equation (4) are time-varying and obtained by the trial-and-error method. Compared to the CCI scheme, the proposed scheme implements inflation once the parameter estimation is turned on, and thus is expected to be more effective in preventing the ensemble spread from a rapid decline and the filter from divergence.

Now, the entire assimilation and inflation process can be summarized as follows. In the assimilation stage, the parameter analyses are obtained by Equations (1) and (2). In the inflation stage, the estimated parameters are adjusted by Equations (3) and (4). Following these steps, a group of comparison experiments are conducted to judge whether the new scheme is better or worse than the previous schemes. Gam2 (see the parameters description in Section 2.4), which is the most sensitive parameter, is estimated with the above six inflation schemes. The initial value is obtained by perturbing gam2 with 20% of its truth.

The convergence time of the parameter estimation, the estimated parameter values, and relative errors are shown in Table 1. The convergence time is defined by the time taken to reach the moment when the difference between the estimated value and the truth value fluctuates in the range of $\pm 10\%$ of the estimated value.

Table 1. Convergence times, the estimated parameter values, and relative errors in parameter estimation with six inflation schemes.

Schemes	FI	CCI	EPES	RTPS	RTPP	New
Convergence times	296	323	27	312	365	288
Estimated values	0.7631	0.7793	0.8006	0.7318	0.7249	0.7492
Relative errors(%)	1.75	3.91	6.75	2.42	3.35	0.11

It can be seen that the new proposed scheme achieves the best estimation. Although the EPES scheme has the shortest convergence time, it converges to the worst estimation. For the other schemes, the new scheme converges the fastest. The results indicate that the new scheme is better than any previous schemes. The comparisons are also implemented in joint estimation of multiple parameters, and a similar result is obtained. Therefore, the new proposed scheme is employed to conduct parameter estimation in this study.

2.4. Model Parameters

Referring to the work of Wu et al. (2016) [12], six parameters in the SST tendency equation are selected, as shown in Table 2. The variation in ocean upwelling and subsurface temperature are controlled by these parameters. Thus, they are very important to the SSTA simulation and prediction. The truths refer to the values in the original LDEO5 model used in Chen et al. (2004) [1]. The initial guesses are obtained by perturbing the parameter with 20% or -20% of its truth.

Table 2. List of parameters (Pars.) in the Z-C model and their truths and initial guesses. Here, h represents the upper layer depth.

Pars.	Physical Meanings	Units	Truths	Initial Guesses
gam1	Strength of mean upwelling advection term	-	0.75	0.6
gam2	Strength of anomalous upwelling advection term	-	0.75	0.6
tda1	Amplitude of subsurface temperature anomaly + h perturbations	$^{\circ}\text{C}$	28	22.4
tda2	Amplitude of subsurface temperature anomaly for $-h$ perturbations	$^{\circ}\text{C}$	-40	-48
tdb1	Affect the nonlinearity of subsurface temperature anomaly for + h perturbations	m^{-1}	1.25	1.0
tdb2	Affect the nonlinearity of subsurface temperature anomaly for $-h$ perturbations	m^{-1}	3.0	2.4

2.5. Experimental Designs

2.5.1. Experimental Designs in OSSE Framework

With the above preparation, parameter estimation and ENSO ensemble experiments are carried out in the OSSE framework to investigate the effectiveness of the parameter estimation based on the LETKF and the impact of the parameter estimation on ENSO prediction.

In the OSSE framework, the truth of observation is generated by running the truth model, in which the parameter values are the truth values. Obviously, the truth of the observation contains information about the states and parameters. The truth model is integrated for 150 years from a prescribed initial condition. The SSTA generated by running the truth model shows the ENSO characteristics, such as the amplitude, period, and asymmetry. Besides generating SSTA observations, the truth model also generates observations for other variables, such as the zonal wind stress anomaly (ZWA), upper layer depth anomaly (ULDA), and subsurface temperature anomaly (TeA), which are used to assess the impact of parameter estimation on the model simulation capabilities.

Considering the estimated parameters are derived from the SST tendency equation in the Z-C model, SSTA data were selected for assimilation. For simplicity, the grids selected from every two grids of the Z-C model were assumed as observation locations. A Gaussian white noise, whose mean and variance were set 0°C and 0.1°C , respectively, was added to the true SSTA at the observation locations. The noise simulates the observational error. The mean and variance of the initial ensemble perturbations of each parameter were 0% and 25% of its initial guess (shown in Table 1), respectively. The monthly data were assimilated into the model once a month, and the ensemble size was 100. Parameter estimation started

after a 5-year only state estimation process, when the model reaches a quasi-equilibrium state [8]. Note that parameter estimation is accompanied by state estimation during the entire assimilation process. For simplicity, we omit the term state estimation hereafter.

To further highlight the effect of parameter estimation, single parameter estimation (SPE) and multiple parameter estimation (MPE) experiments were designed. According to the sensitivity study, $\text{gam}2$ is the most sensitive parameter. Thus, it was selected in the SPE experiment. For comparison, an experiment with only state estimation (onlySE) was also carried out. Thus, there are three groups of assimilation experiments, as shown in Table 3. According to a trial-and-error process, the values of a and b in Equation (4) are also listed in Table 3.

Table 3. Experimental designs in the OSSE framework. Note that six parameters are estimated in MPE, so there are six pairs of values of a and b in MPE.

Experiments	Assimilated	Estimated	(a,b)
onlySE	SSTA	SSTA	(-, -)
SPE	SSTA	$\text{gam}2 \backslash \text{SSTA}$	(0.15, 0.08)
MPE	SSTA	$\text{gam}1 \backslash \text{gam}2 \backslash \text{tda}1 \backslash \text{tda}2 \backslash \text{tdb}1 \backslash \text{tdb}2 \backslash \text{SSTA}$	(0.31, 0.016) \ (0.30, 0.015) \ (4.12, 1.69) \ (6.58, 3.44) \ (0.09, 0.05) \ (0.45, 0.20)

Accordingly, there are three prediction experiments: the first one is initialized by the analysis of onlySE and uses an initial guess of the model parameters, and the other two are initialized by the analysis of the joint state-parameter estimate (SPE and MPE, respectively) and use the parameter values averaged over the last 100 years, which will be further explained. All prediction experiments were carried out for a period of 150 years, making a monthly forecast, and each forecast lasting for 12 months.

In a noise-free model, stochastic optimals (SOs) [44] are used to capture the effects of stochastic atmospheric processes on model prediction errors. The definitions of SOs followed that in Gao et al. (2020) [3]. For white noise across the interval of time, the SOs are eigenvectors of the operator S , which is expressed as follows:

$$S = \int_0^T R^*(0, t) R(0, t) dt, \quad (5)$$

where T is the forecast interval of interest, $R(0, t)$ is the forward tangent propagator of the linearized dynamical model that advances the state vector of the system from time 0 to t , and R^* is the adjoint of R . In this study, the SOs are added to the surface wind anomalies during the entire prediction integration. With the stochastic optimal perturbations, 100 ensemble forecasts were generated in each experiment.

2.5.2. Experimental Designs in the Real World

Since our ultimate purpose of developing a new inflation scheme for parameter estimation is to study the real world, real-world experiments were carried out in this section.

In real-world experiments, parameter estimation and ENSO ensemble prediction experiments were carried out from January 1970 to December 2000. The assimilated SSTA data were obtained from the latest version of the ERSST dataset (ERSST V5, <https://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v5>, accessed on 18 September 2020). In addition, the relevant SODA reanalysis data (<http://iridl.ldeo.columbia.edu/SOURCES/.CARTON-GIESE/.SODA/.v2p0p2-4/>, accessed on 18 September 2020) were compared with the model results, assessing the simulation capability of the state variables. The other settings, including but not limited to assimilation frequency, ensemble size, and forecast duration, are the same as that in Section 2.5.1.

In theory, MPE is more effective than SPE since it is expected to further reduce model errors and improve model simulation and prediction capabilities. Therefore, only MPE and the corresponding ENSO ensemble prediction experiments were performed in this section. Additionally, as a control experiment, onlySE is carried out. The two experimental designs are listed in Table 4. According a trial-and-error process, the values of a and b in Equation (4) are also listed in Table 4.

Table 4. Experimental designs in the real world. Note that six parameters are estimated in MPE, so there are six pairs of values of a and b in MPE.

Experiments	Assimilated	Estimated	(a,b)
OnlySE	SSTA	SSTA	(-, -)
MPE	SSTA	gam1\gam2\tda1\ tda2\tdb1\tdb2\SSTA	(0.02,0.01)\(0.02,0.01)\ (0.80,0.38)\(1.14,0.54)\ (0.04,0.02)\(0.08,0.04)

It is expected that correlations between any given observation and variable to be estimated decrease with distance in the physical world. Therefore, correlations between observations and variables that are spatially remote may be regarded as spurious. The spurious correlations degrade the quality of the analysis. In order to remove the spurious correlations between the model variables and remote observations, the Gaspri–Cohn function (GC function) [3,45], which is the most widely used function to cut off the long-distance correlations, is used in this study. In this function, the correlation coefficient is 1 when the distance is zero. The coefficient decreases gradually as the distance increases, until it is 0, where the distance is larger than the de-correlation length. After a trial-and-error process, the best result is obtained when the de-correlation length is four times as much as the distance between the adjacent grids.

3. Results

With the preparations and experimental designs in Section 2, the results in the OSSE framework and real-world experiments are reported in this section. To further verify the optimized parameters in the real-world experiment, the validation results are given.

3.1. Results of the OSSE Framework

A successful parameter estimation should gradually adjust the ensemble mean of the parameter from the initial guess to the truth with assimilation. Figure 2 shows the temporal evolution of the estimated gam2, its root mean square error (RMSE) relative to observations, and the ensemble spread with assimilation length in SPE. It is evident that the estimated gam2 converges to a constant value close to the truth of 0.75 after about four decades, in spite of a large initial error. Thus, the estimated parameter value is obtained using the average of the value over the last 100 years, labeled in subplot subfigure (a). As can be seen in subfigure (b), the ensemble spread tends to be stable after decreasing by 2 orders of magnitude, indicating that the ensemble members still contain definite divergence; although the RMSE has an initial shock, it is still very close to the spread after four decades, indicating the success of SPE with the proposed inflation scheme. To examine the validity of the inflation scheme in other parameter estimations, gam1, tda1, tda2, tdb1, and tdb2 were also estimated. Similar results were obtained.

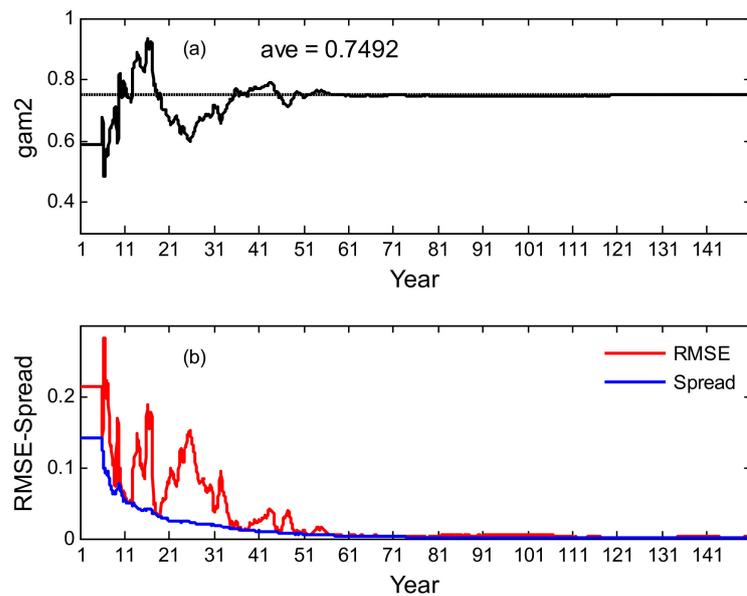


Figure 2. Temporal evolution of (a) estimated gam2, and (b) its RMSE relative to the “observations” and ensemble spread with assimilation length in SPE: bold solid line indicates the ensemble mean of gam2 and the dashed line is the truth in panel (a).

Figure 3 shows the estimated values of the joint estimation of six parameters in MPE. It is noted that all parameters converge to the respective truths after about four decades of adjustment. Specifically, the estimated gam1, tda1, and tdb1 are very close to their truths, while the estimated gam2, tda2, and tdb2 are much closer to the truths than their initial guesses, although they are still not as precise as the truths. Similar to Figure 2a, the estimated parameter values are obtained using the average of the values over the last 100 years, labeled in panels (a) to (f) in Figure 3. These values are used to perform the prediction experiments.

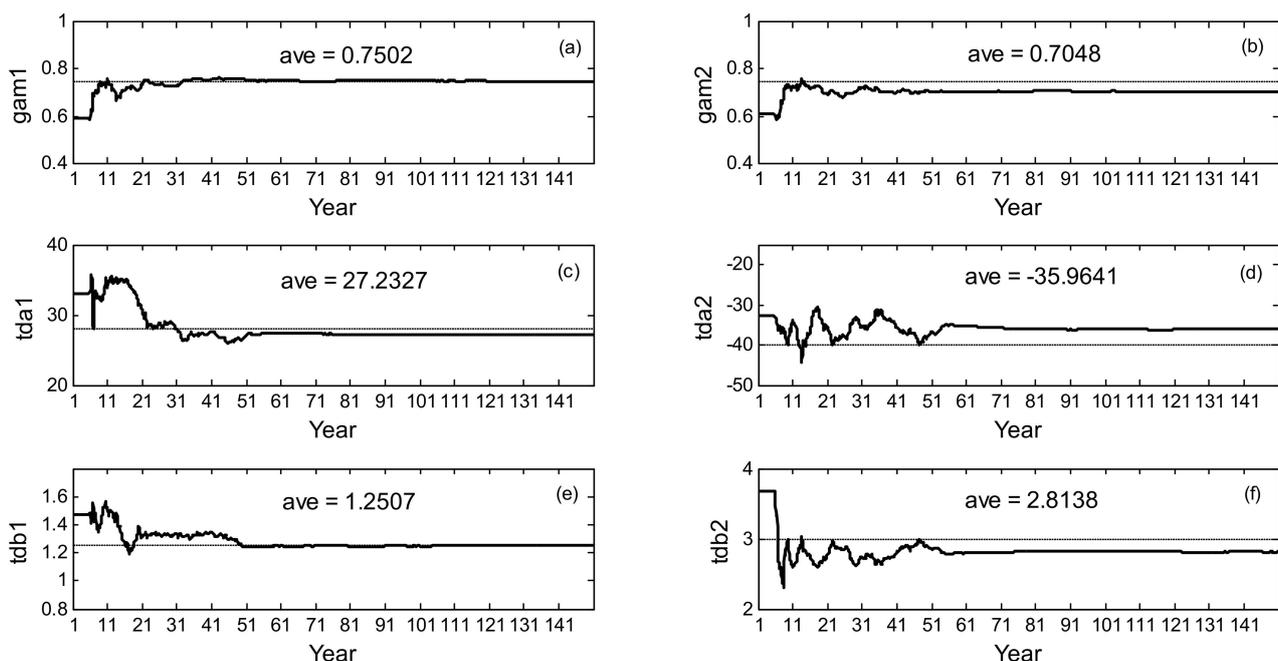


Figure 3. Temporal evolution of six estimated parameters with assimilation length in MPE: (a–f) the six subfigures representing gam1, gam2, tda1, tda2, tdb1, and tdb2, respectively; bold solid lines indicate the ensemble mean of the parameters, and the dashed lines are their truths.

The relative errors of the initial and estimated parameters against the true values are listed in Table 5. It can be seen that the relative errors of the estimated values decrease significantly when compared to the initial values, indicating successful results of MPE with the proposed inflation scheme.

Table 5. Relative errors of the initial and estimated parameters against the true values.

Relative Errors	gam1	gam2	tda1	tda2	tdb1	tdb2
Initial	−20%	−20%	20%	20%	20%	20%
Estimated	0.03%	−6%	−2.7%	10.1%	0.05%	−6.2%

The improvement of the state variables in conjunction with parameter estimation is also examined. Figure 4 shows the temporal evolution of RMSE between the simulated state variables and the observed counterparts during the last 100 years in three experiments listed in Table 3. The time averaged RMSE for each experiment is labeled in panels (a) to (d) in Figure 4. It should be mentioned that all RMSEs in the three experiments decrease by one order of magnitude compared with their initial RMSEs, which are 0.4587, 0.3307, 17.3979, and 1.3690, respectively (not shown here). The figure shows that for the four state variables, the MPE gets the smallest RMSE, followed by SPE, and onlySE gets the largest RMSE, showing the high effectiveness in parameter estimation by the newly proposed scheme in improving the state analysis. Since the assimilated data is SSTA, the simulation of SSTA is improved most significantly; other state variables are also improved to some extent through the dynamic model within the framework of coupling assimilation.

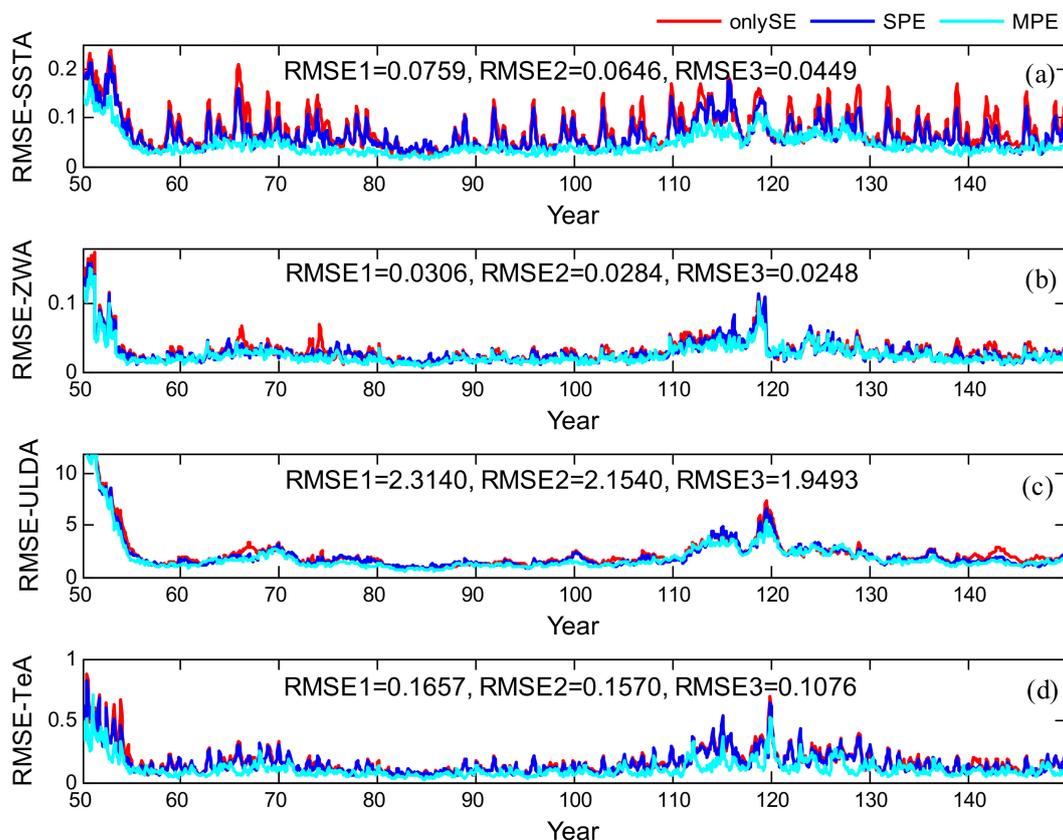


Figure 4. Time series of RMSE between the simulated state variables and the corresponding “observed” values during the last 100 years in three experiments: From (a) to (d), the four subfigures represent RMSEs of SSTA, ZWA, ULDA, and TeA, respectively.

The anomaly correlation coefficient (ACC) of the predicted SST anomalies against the true counterparts in Niño3.4 region (shown in Figure 1) is used to evaluate the predictability of ENSO. Because the parameter values used in the three prediction experiments are obtained by the averages of the values over the last 100 years, Figure 5 shows the variation in ACCs with respect to lead time (in months) and start month during the last 100 years. It is shown that MPE obtains the highest prediction skill, SPE gets a slightly worse skill, and onlySE gets the worst skill. This result is not surprising since MPE gives the best parameters and state estimates and onlySE uses biased parameters.

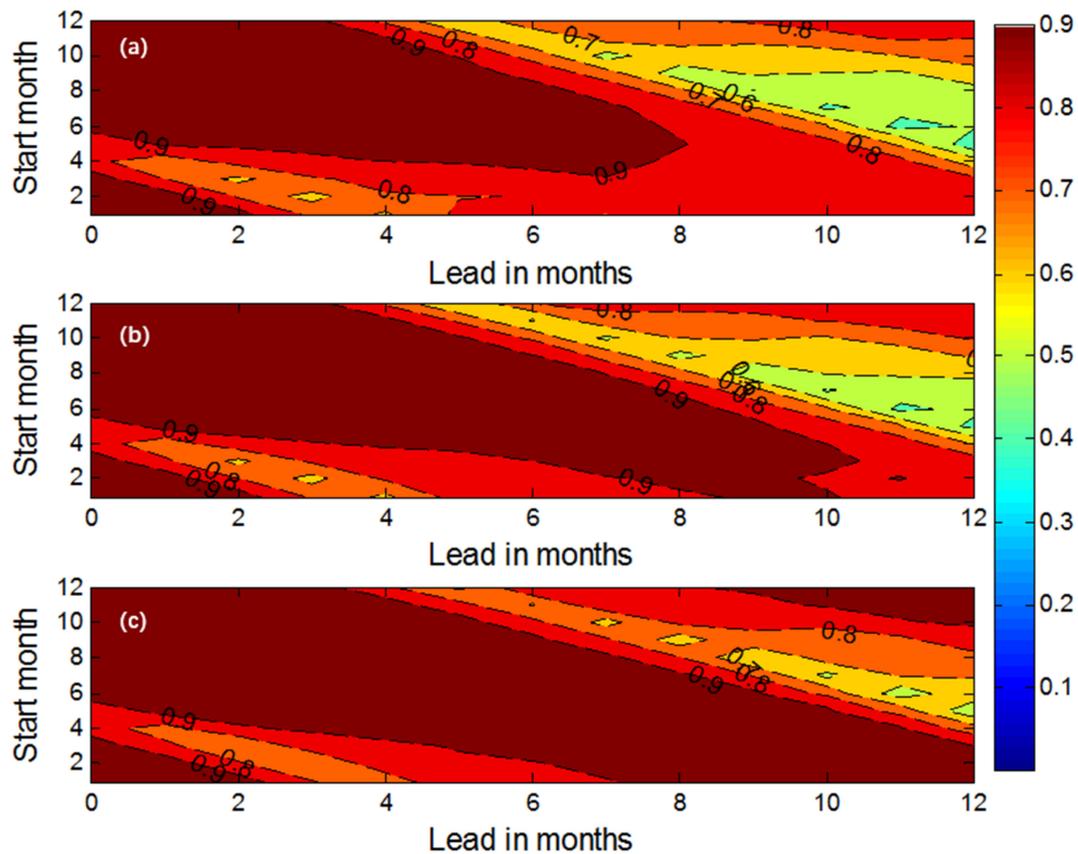


Figure 5. Variation in ACCs averaged over the last 100 years in the three forecasting experiments: (a) onlySE, (b) SPE, and (c) MPE.

The spring predictability barrier (SPB) is a well-known feature of ENSO forecasts, which refers to the phenomenon that the prediction skills of most ENSO prediction models tend to experience a significant decline in April and May (Figure 5) [46,47]. Although considerable progress has been made in investigating the SPB [48–50], its cause remains elusive. Many studies have shown that the SPB is an inherent feature of the ENSO, whereas Chen et al. (2004) [1] suggested that improving initialization could effectively reduce the predictability barrier. Interestingly, Figure 5 shows that compared with SPE and onlySE, MPE alleviates SPB to a certain extent, suggesting that model error also should have an influence on SPB.

The above results showed that the erroneously set parameter values limited the ENSO prediction skill. Parameter estimation with the proposed inflation scheme could reduce the model error and improve the ENSO prediction skill in the Z-C model. More specifically, MPE achieved greater state simulations and ENSO prediction skills than SPE. It is worth mentioning that we found that model error is also one of the reasons for SPB in an ENSO prediction model.

3.2. Results in Real-World Experiments

Figure 6 shows temporal evolution of the six estimated parameters with assimilation length in MPE. All six parameters are relatively stable after years of initial shocks. Following the methodology in Section 3, the estimated parameter values (labeled in subfigures) were obtained using the average of values over the last 20 years. The estimates are close to the respective default values of the model, indicating that the parameter estimation is generally reasonable. However, Hansen and Penland (2007) [30] pointed out that in the case of real-world assimilation, a model parameter may not even have a truth value, or it may not converge to a constant because the model parameter may be heavily dependent on state variables. Therefore, the concept that parameter estimation converges to a certain deterministic value should not be expected. On the contrary, parameter values should vary over time (and space) in order to reflect their trait of being state dependent. The relatively large variation in the estimated parameters, which occurred around 1982/83 and 1997/98 as a result of the two super El-Niño events in Figure 6, further illustrate this concept. This partially explains why climate models whose parameters are never updated often yield poor skills in forecasting ENSO for some periods of time. It should be noted that the variation in the parameters seem to contradict with the conclusion of Zhao et al. (2019) [25]. They suggested that the parameter convergence is robust and tends to be consistent, whereas our results show that the parameters may change as the mean states, even instant states, vary.

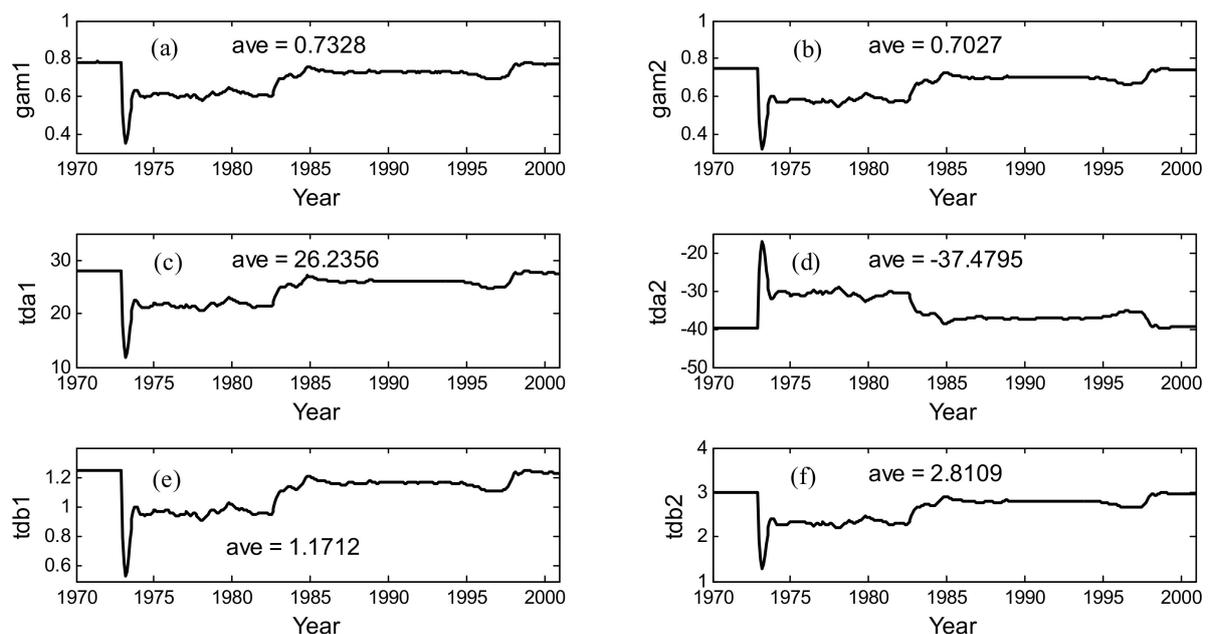


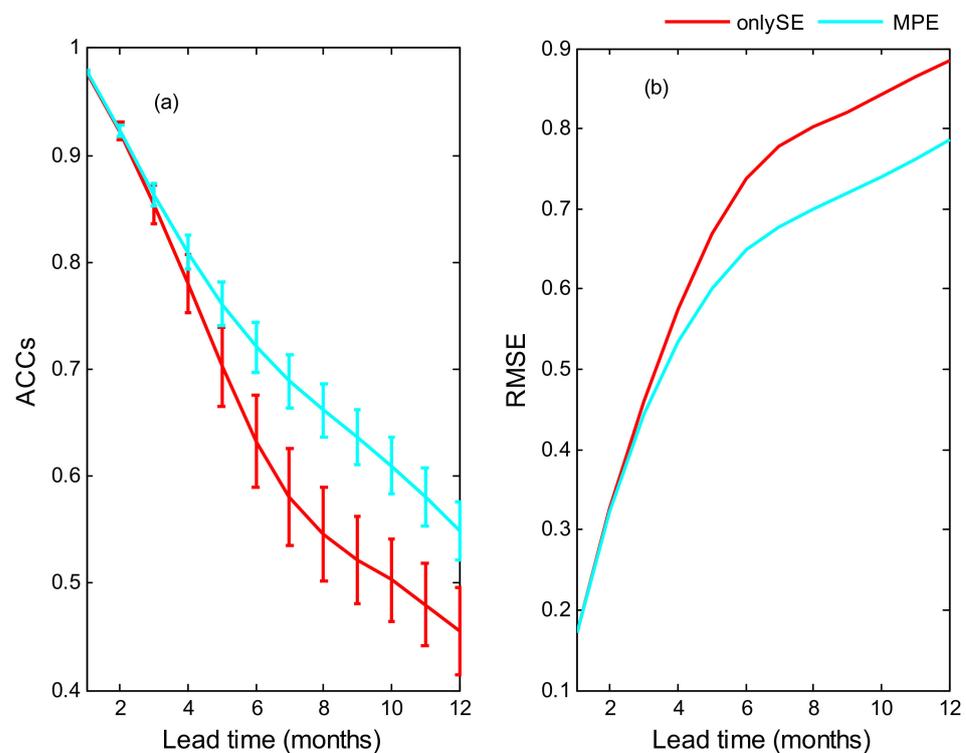
Figure 6. Temporal evolution of the six estimated parameters with assimilation length in MPE: (a–f) the six subfigures represent gam1, gam2, tda1, tda2, tdb1, and tdb2, respectively.

The simulated SSTA, ZWA, ULDA, and TeA in MPE and onlySE were compared with the corresponding values in the SODA reanalysis dataset, respectively. Table 6 shows the correlation coefficients and RMSEs of the averaged four state variables over Niño3.4 region during the last 20 years (i.e., from January 1981 to December 2000), respectively. It is shown that all correlation coefficients in MPE are higher than those in onlySE, while all RMSEs in MPE are lower than those in onlySE. All correlation coefficients are statistically significant at the 99% confidence level. The results verify the conclusion that parameter estimation reduces the model error.

Table 6. Averaged correlation coefficients and RMSEs of the simulated SSTA, ZWA, ULDA, and TeA in two experiments against the corresponding counterparts in the SODA reanalysis dataset.

State Variables	SSTA		ZWA		ULDA		TeA	
Exps.	onlySE	MPE	onlySE	MPE	onlySE	MPE	onlySE	MPE
Correlation coefficients	0.9256	0.9594	0.8248	0.8356	0.7172	0.8205	0.7605	0.7893
RMSEs	0.3071	0.2818	0.2318	0.2271	6.6496	6.5208	0.9197	0.9092

Figure 7 shows the ACCs and RMSEs of the predicted SST anomalies based on onlySE and MPE against the true counterparts in the Niño3.4 region. It is shown that the prediction skill of MPE is significantly higher than that of onlySE. As described in the OSSE framework, MPE provides better initial conditions and more reasonable parameter values, and it improves the ENSO prediction skill of the model. To confirm the significance of the skill differences between the two experiments, the error bars are shown in panel (a), which indicates that the differences exceeded the sampling error beyond the 5-month leads.

**Figure 7.** The (a) ACCs and (b) RMSEs of the predicted SST anomalies against the true counterparts in the Niño3.4 region during the period 1981–2000.

The observed and predicted SST anomalies averaged over the Niño3.4 region based on MPE are shown in Figure 8. The forecasts fit the observations sufficiently, which indicates that the model is able to predict most of the warm and cold events that occurred during 1981–2000, especially the amplitudes of relatively large El Niños and La Niñas. Over a lead time up to 3 months, the correlation skill is about 0.93; and over a lead time up to 6 months, the correlation skill is about 0.76. These skills are comparable with the current best skills of advanced coupled models.

To further evaluate the prediction of warm and cold events, an El Niño is defined when the Niño3.4 index is greater than $1\text{ }^{\circ}\text{C}$, whereas a La Niña is defined when the Niño3.4 index is less than $-1\text{ }^{\circ}\text{C}$, which refers to Chen et al. (2004) [1]. Composites of these observed El Niños and La Niñas are displayed in Figure 9, together with the corresponding prediction composites at the leading 3 months based on onlySE and MPE. In general, the

model well captures the spatial patterns of both El Niño and La Niña. However, there are still some differences in amplitude. Compared to onlySE, MPE improves the El Niño predictions in the central and eastern Pacific Oceans, but is not effective for La Niña predictions. Figure 6 showed that the super El Niños that occurred in 1982/83 and 1997/98 affected the parameter estimation; in turn, parameter estimation improves the prediction of warm events, as shown in Figure 9. For La Niñas, the relationship is not obvious. The reason for this phenomenon may be that the intensity of a La Niña is relatively small, resulting in difficulties in improving the parameter estimation and forecasting skill.

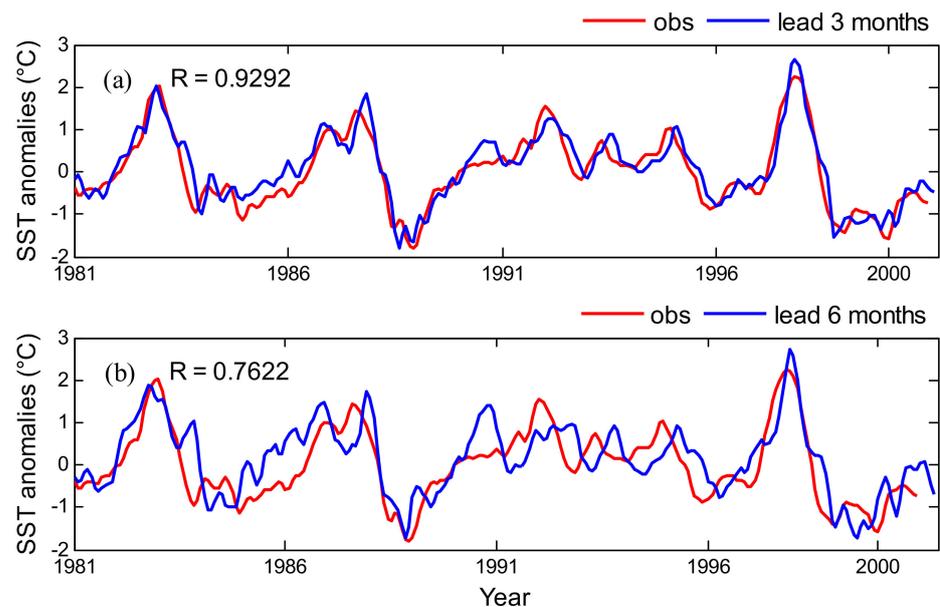


Figure 8. Retrospective predictions of El Niños and La Niñas during 1981–2000: the red curve is the observed values and the blue curves are the predicted values at a 3-month and 6-month lead in (a) and (b) subfigures, respectively.

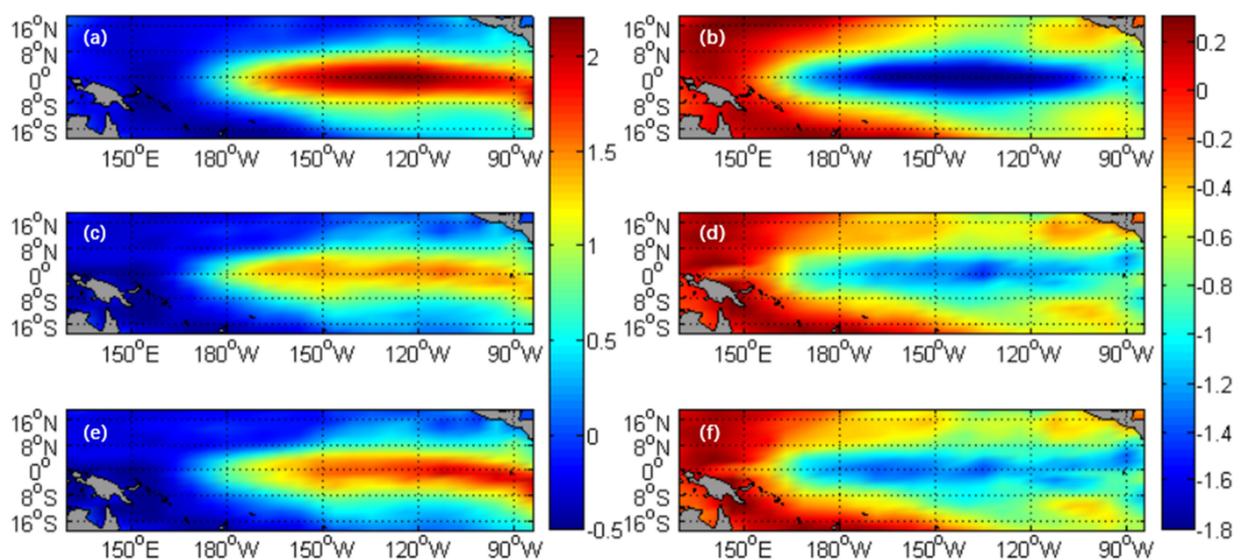


Figure 9. Composites of El Niños and La Niñas: (a,b) represent composites of observations, (c,d) represent predictions in onlySE, and (e,f) represent predictions in MPE.

The above results showed that, in the real-world scenario, the model parameters converged to new optimal values and the optimized parameters reduced the model error

and thus improved the ENSO prediction skill, especially for skill with a long lead time. Specifically, parameter estimation was influenced by super El Niños, and as a result it could improve the prediction of warm events.

3.3. Validation of Optimized Parameters

Although the above results indicated that the optimized parameters improved the ENSO prediction skill, the parameter values were estimated by assimilating the observations during the entire period of 1981–2000, which may bring artificial skills in prediction verification (shown in Figure 6). For further verification, we conducted a group of comparative experiments during the period 2001–2020. The first group employs the optimized parameter values, and the other group of experiments utilizes the default parameter values as noted in the Z-C model. The settings of the two experiments are listed in Table 7.

Table 7. Settings of the two experiments during the validation period of 2001–2020.

Experiments	Parameter Values (Truncated to Two Decimal Places)
Optimized Pars	gam1 = 0.73, gam2 = 0.70, tda1 = 26.24, tda2 = −37.48, tdb1 = 1.17, tdb2 = 2.81
Default Pars	gam1 = 0.75, gam2 = 0.75, tda1 = 28.00, tda2 = −40.00, tdb1 = 1.25, tdb2 = 3.00

Figure 10 shows the ACCs and RMSEs of the predicted SST anomalies against the true counterparts in the Niño3.4 region for the two experiments. When compared to the parameters in the truth model, the optimized parameters improve the ENSO prediction skills for predictions beyond 5–6-month leads, which demonstrates clearly the advantage of parameter estimations in climate prediction. To confirm the significance of the skill differences between the two experiments, the error bars are shown in panel (a), which indicates that the differences exceed the sampling error beyond the 6-month leads.

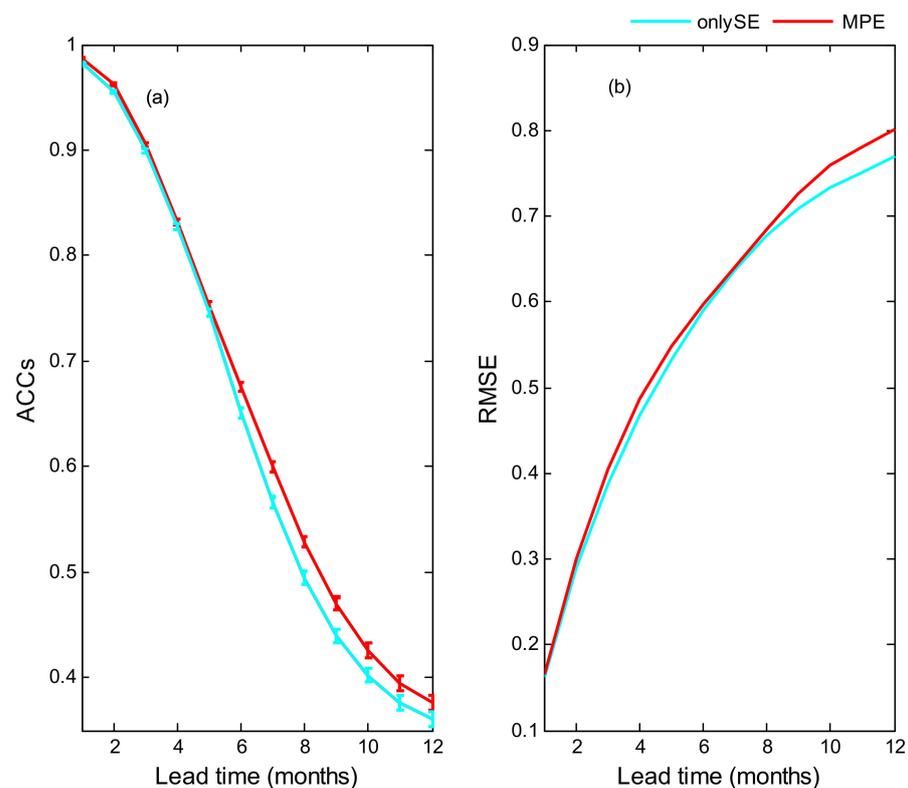


Figure 10. The (a) ACCs and (b) RMSEs of the predicted SST anomalies against the true counterparts in the Niño3.4 region during the period 2001–2020.

4. Discussion

A large number of studies in perfect scenarios have shown that parameter estimation is of great significance to improve the simulation and prediction capability of the model. Due to the constant or slow-varying characteristics of parameters, parameter estimation in the real world is an urgent and challenging research aim. However, few studies have addressed this problem. Some studies have shown that the covariance inflation scheme could improve the accuracy of parameter estimation. This study proposed a new covariance inflation scheme based on LETKF and employed it in parameter estimation in the Z-C model.

The parameter values in Z-C model were determined several decades ago. In recent decades, the oceanic background circulation and the physical mechanism of ENSO have changed greatly [51–53]; thus, the parameter values are no longer suitable for the present scenario. Our results confirmed this concept. Note that the prediction skill was slightly improved but not significantly improved during the test period (see Figure 10). There are many factors leading to this result, such as only SSTA were assimilated, and the stochastic atmospheric processes were not precisely simulated.

As pointed out in this study, the parameters that depend heavily on the state variables should vary over time (and space). In the real-world scenario, factors affecting the parameters are very complex. For example, strong El Niño events are often accompanied by atmospheric random processes (such as westerly wind bursts). However, the Z-C model still has defects in the expression of these factors, which makes parameter estimation more difficult.

In general, it is interesting to apply multiple source data assimilation and westerly wind bursts for further exploring ENSO predictability in the future.

5. Conclusions

In this study, we proposed a new inflation scheme based on LETKF and assessed its application in parameter estimation and ENSO prediction based on the Z-C model in the OSSE framework and real-world experiments, respectively.

Firstly, parameter estimation and ENSO ensemble prediction experiments were carried out within the OSSE framework. The results showed that both SPE and MPE made the estimated parameters converge to the truths. The simulations of the state variables indicated that parameter estimation reduced the model error, and MPE further reduced the model error than SPE. By providing better initial conditions and parameter values, MPE improved the ENSO prediction skill and alleviated the SPB phenomenon. Secondly, parameter estimation and ENSO ensemble prediction experiments were carried out in real-world experiments. The results showed that the estimated parameters converged to new optimal values except for two abnormal oscillations during the two super El Niño events, indicating that the model parameter depends heavily on the state variable. The optimized parameters improved the ENSO prediction significantly, primarily in prediction of warm events. In addition, the optimized parameters were validated by comparative experiments during a test period that is independent of the parameter. The results indicated that the optimized parameters improved the prediction ability of the Z-C model. In general, this study provided an effective parameter estimation strategy to improve climate models and further climate predictions in the real world.

Author Contributions: Conceptualization, Y.G. and Y.T.; formal analysis, Y.G. and X.S.; funding acquisition, Y.T.; methodology, Y.G., Y.T. and Z.S.; software, Y.G. and X.S.; validation, Y.G. and Y.T.; writing—original draft, Y.G.; writing—review and editing, Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported jointly by the National Key R&D Program of China (Grant no. 2017YFA0604202), National Natural Science Foundation of China (Grant nos. 41690124 and 41690120), Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai, Grant no. 311021009), and Oceanic Sustainability based Marine Science and Technology Cooperation in Maritime Silk Road and Island Countries.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The SSTA in this study are available from the latest version of the ERSST dataset (ERSST V5, <https://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v5>, accessed on 18 September 2020). The ZWA, ULDA, TeA datasets are available from SODA reanalysis data (<http://iridl.ldeo.columbia.edu/SOURCES/.CARTON-GIESE/.SODA/.v2p0p2-4/>, accessed on 18 September 2020).

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

Appendix A

The state vector augmentation technique is a useful tool to conduct parameter estimation. The idea of the technique is borrowed from strong coupled data assimilation. Assume the model is composed of several parts, for simplicity, oceanic component and atmospheric component. If there are observations from the two components, strongly coupled based on LETKF are conducted as follows:

Step (1). The i th member of \tilde{H} can be calculated by

$$\tilde{H}_i = Hx_i - \overline{Hx}_i = \begin{bmatrix} H_\eta \eta_i - \overline{H_\eta \eta}_i \\ H_\zeta \zeta_i - \overline{H_\zeta \zeta}_i \end{bmatrix}, \quad (\text{A1})$$

where η and ζ represent oceanic and atmospheric components, respectively.

Step (2). The observation error covariance matrix in Step (2) are correspondingly written as

$$R = \begin{bmatrix} R_\eta & R_{\eta\zeta} \\ R_{\eta\zeta}^T & R_\zeta \end{bmatrix}. \quad (\text{A2})$$

Note that, it is generally believed that the observations from different components do not affect each other. Thus, $R_{\eta\zeta}$ in Equation (A2) is zero. Based on this, the projection of the analysis error covariance matrix into the ensemble space is calculated, expressed as $\tilde{P}^a = [\tilde{H}^T R^{-1} \tilde{H} + (Ne - 1)I]^{-1}$.

Step 3: The i th perturbation of the predictions X^b_i can be written as

$$X^b_i = \begin{bmatrix} \eta^b_i - \overline{\eta^b}_i \\ \zeta^b_i - \overline{\zeta^b}_i \end{bmatrix}, \quad (\text{A3})$$

Based on this, the Kalman gain is calculated, expressed as $K = X^b \tilde{P}^a \tilde{H}^T R^{-1}$.

Step 4: The background mean \bar{x}^b can be written as

$$\bar{x}^b = \begin{bmatrix} \overline{\eta^b} \\ \overline{\zeta^b} \end{bmatrix}. \quad (\text{A4})$$

With equations displayed above, the analysis mean \bar{x}^a and perturbation X^a can be calculated using Equations (1) and (2), which are not shown here.

Step 5: The analysis is obtained by adding the analysis perturbation to the mean, expressed as

$$x^a = \bar{x}^a + X^a. \quad (\text{A5})$$

According to the above steps, observations in one component can be used to estimate variables in the other component. Parameters can be recognized as one of the components in a coupled model. Since there is no direct observation of parameters in real scenarios, we can omit the variables in parameter component, for example, variables relative to ζ in Equations (A1) and (A2). However, the variables relative to ζ in (A3) and (A4) should not be omitted.

Appendix B

According to different inflation factors, i.e., γ in Equation (3), the previous inflation schemes are introduced as below:

FI scheme:

The simplest scheme is to set γ to a fixed inflation (FI) factor that is slightly larger than 1 (Anderson, 2001; Ruiz et al., 2013a, b; Luo et al., 2014; Han et al., 2015; O’Kane et al., 2019).

CCI scheme:

In the CCI scheme, the inflation factor in Equation (3) is expressed as

$$\gamma = \begin{cases} 1, & \sigma_{\beta} \geq b \\ \frac{b}{\sigma_{\beta}}, & \sigma_{\beta} < b \end{cases}, \quad (\text{A6})$$

where σ_{β} represents the posterior standard deviation of the parameter ensemble, and b is the threshold (Tong and Xue, 2008a, b; Han et al., 2015; Aksoy et al., 2006a, b; Wu et al., 2012a, 2013; Zhao et al., 2019).

EPES scheme:

In the EPES scheme, the inflation factor in Equation (3) can be written as

$$\gamma = \lambda \sqrt{\frac{\text{Ne}}{(\text{Ne} - 1)\text{tr}(\tilde{\mathbf{P}}^a)}}, \quad (\text{A7})$$

where $\tilde{\mathbf{P}}^a$ is the projection of the analysis error covariance matrix into the ensemble space, tr stands for the matrix trace, Ne is the ensemble number, and λ is a multiplying factor that is slightly less than 1 (Ruiz et al., 2013).

RTPP scheme:

In the RTPP scheme, the inflated ensemble can be expressed as

$$\tilde{\beta}_i = \bar{\beta}^a + (1 - \alpha)(\beta_i^a - \bar{\beta}^a) + \alpha(\beta_i^b - \bar{\beta}^b) \quad (\text{A8})$$

where β_i^a and β_i^b denote the posterior and prior parameters, and $\bar{\beta}^a$ and $\bar{\beta}^b$ denote the ensemble means of β_i^a and β_i^b , respectively. α is a relaxation factor between 0 and 1 (Zhang et al., 2004; Whitaker and Hamill, 2012).

RTPS scheme:

In the RTPS scheme, the inflation factor in Equation (3) can be written as

$$\gamma = \left(\alpha \frac{\sigma^b - \sigma^a}{\sigma^a} + 1 \right), \quad (\text{A9})$$

where σ^b and σ^a denote the prior and posterior ensemble standard deviation at each analysis grid, respectively, and α is a relaxation factor between 0 and 1, but different from that in the RTPP algorithm (Whitaker and Hamill, 2012).

References

- Chen, D.K.; Cane, M.A.; Kaplan, A.; Zebiak, S.; Huang, D. Predictability of El Niño over the past 148 years. *Nature* **2004**, *428*, 733–736. [[CrossRef](#)]
- O’Kane, T.J.; Sandery, P.A.; Monselesan, D.P.; Sakov, P.; Stevens, L. Coupled Data Assimilation and Ensemble Initialization with Application to Multiyear ENSO Prediction. *J. Clim.* **2019**, *32*, 997–1024. [[CrossRef](#)]
- Gao, Y.Q.; Liu, T.; Song, X.S.; Shen, Z.Q.; Tang, Y.M.; Chen, D.K. An extension of LDEO5 model for ENSO ensemble predictions. *Clim. Dyn.* **2020**, *55*, 2979–2991. [[CrossRef](#)]
- Latif, M.; Sperber, K.; Arblaster, J.; Braconnot, P.; Chen, D.; Colman, A.; Cubasch, U.; Cooper, C.; Delecluse, P.; DeWitt, D.; et al. ENSIP: The El Niño simulation intercomparison project. *Clim. Dyn.* **2001**, *18*, 255–276. [[CrossRef](#)]
- Stainforth, D.A.; Aina, T.; Christensen, C.; Collins, M.; Faull, N.; Frame, D.J.; Kettleborough, J.A.; Knight, S.; Martin, A.; Murphy, J.M. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **2005**, *433*, 403–406. [[CrossRef](#)] [[PubMed](#)]

6. Zheng, F.; Wang, H.; Zhu, J. ENSO ensemble prediction: Initial error perturbations vs. model error perturbations. *Chin. Sci. Bull.* **2009**, *54*, 2516–2523. [[CrossRef](#)]
7. Tao, L.J.; Gao, C.; Zhang, R.H. Model parameter-related optimal perturbations and their contributions to El Niño prediction errors. *Clim. Dyn.* **2019**, *52*, 1425–1441. [[CrossRef](#)]
8. Zhang, S.Q.; Liu, Z.Y.; Rosati, A.; Delworth, T. A study of enhance parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus A* **2012**, *64*, 10963. [[CrossRef](#)]
9. Qi, Q.Q.; Duan, W.S.; Zheng, F.; Tang, Y.M. On the “spring predictability barrier” for strong El Niño events as derived from an intermediate coupled model ensemble prediction system. *Sci. China Earth Sci.* **2017**, *60*, 1614–1631. [[CrossRef](#)]
10. Song, J.Q.; Cao, X.Q.; Zhang, W.M.; Zhu, X.Q. Estimating parameters for coupled air-sea model with variational method. *Acta Phys. Sin.* **2012**, *61*, 110401. (In Chinese) [[CrossRef](#)]
11. Wu, X.; Zhang, S.; Liu, Z.; Rosati, A.; Delworth, T.L.; Liu, Y. Impact of geographic-dependent parameter optimization on climate estimation and prediction: Simulation with an intermediate coupled model. *Mon. Weather Rev.* **2012**, *140*, 3956–3971. [[CrossRef](#)]
12. Wu, X.; Han, G.; Zhang, S.; Liu, Z. A study of the impact of parameter optimization on ENSO predictability with an intermediate coupled model. *Clim. Dyn.* **2016**, *46*, 711–727. [[CrossRef](#)]
13. Navon, I.M. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Ocean.* **1998**, *27*, 55–79. [[CrossRef](#)]
14. Annan, J.D.; Hargreaves, J.C.; Edwards, N.R.; Marsh, R. Parameter estimation in an intermediate complexity Earth System Model using an ensemble Kalman filter. *Ocean Model.* **2004**, *8*, 135–154. [[CrossRef](#)]
15. Annan, J.D. Parameter estimation using chaotic time series. *Tellus A* **2005**, *57*, 709–714. [[CrossRef](#)]
16. Du, H.; Huang, S.; Cai, Q.; Cheng, L. Studies of variational assimilation for the inversion of the coupled air-sea model. *Mar. Sci. Bull.* **2009**, *11*, 13–22.
17. Hu, X.M.; Zhang, F.; Nielsen-Gammon, J.W. Ensemble based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study. *Geophys. Res. Lett.* **2010**, *37*, L08802. [[CrossRef](#)]
18. Ito, K.; Ishikawa, Y.; Awaji, T. Specifying air-sea exchange coefficients in the high-wind regime of a mature tropical cyclone by an adjoint data assimilation method. *Sola* **2010**, *6*, 13–16. [[CrossRef](#)]
19. Gao, Y.Q.; Cao, A.Z.; Chen, H.B.; Lv, X.Q. Estimation of bottom friction coefficients based on an isopycnic-coordinate internal tidal model with adjoint method. *Math. Probl. Eng.* **2013**, *2013*, 532814. [[CrossRef](#)]
20. Peng, S.Q.; Li, Y.N.; Xie, L. Adjusting the wind stress drag coefficient in storm surge forecasting using an adjoint technique. *J. Atmos. Ocean. Technol.* **2013**, *30*, 590–608. [[CrossRef](#)]
21. Han, G.; Wu, X.; Zhang, S.; Liu, Z.; Navon, I.M.; Li, W. A study of coupling parameter estimation implemented by 4D-Var and EnKF with a simple coupled system. *Adv. Meteorol.* **2015**, *2015*, 530764. [[CrossRef](#)]
22. Gao, C.; Zhang, R.H.; Wu, X.R.; Sun, J.C. 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](#)]
23. Zhang, J.C.; Li, G.Q.; Yi, J.C.; Gao, Y.Q.; Cao, A.Z. A method on estimating time-varying vertical eddy viscosity for an Ekman layer model with data assimilation. *J. Atmos. Ocean. Technol.* **2019**, *36*, 1789–1812. [[CrossRef](#)]
24. Kang, J.; Kalnay, E.; Liu, J.; Fung, I.; Miyoshi, T.; Ide, K. Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J. Geophys. Res.* **2012**, *117*, D24101. [[CrossRef](#)]
25. Zhao, Y.; Liu, Z.; Zheng, F.; Jin, Y. Parameter optimization for real-world ENSO forecast in an intermediate coupled model. *Mon. Weather Rev.* **2019**, *147*, 1429–1445. [[CrossRef](#)]
26. Li, S.; Zhang, S.; Liu, Z.; Lu, L.; Zhu, J.; Zhang, X.; Wu, X.; Zhao, M.; Vecchi, G.A.; Zhang, R.; et al. Estimating convection parameters in the GFDL CM2.1 model using ensemble data assimilation. *J. Adv. Model. Earth Syst.* **2018**, *10*, 989–1010. [[CrossRef](#)]
27. Shen, Z.Q.; Tang, Y.M.; Gao, Y.Q. The theoretical framework of the ensemble-based data assimilation method and its prospect in oceanic data assimilation. *Haiyang Xuebao* **2016**, *38*, 1–14. (In Chinese) [[CrossRef](#)]
28. Zhang, X.; Zhang, S.; Liu, Z.; Wu, X.; Han, G. Correction of biased climate simulated by biased physics through parameter estimation in an intermediate coupled model. *Clim. Dyn.* **2015**, *47*, 1899–1912. [[CrossRef](#)]
29. Anderson, J.L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **2001**, *129*, 2884–2903. [[CrossRef](#)]
30. Hansen, J.; Penland, C. On stochastic parameter estimation using data assimilation. *Phys. D* **2007**, *230*, 88–98. [[CrossRef](#)]
31. Han, G.; Wu, X.; Zhang, S.; Liu, Z.; Li, W. Error covariance estimation for coupled data assimilation using a Lorenz atmosphere and a simple pycnocline ocean model. *J. Clim.* **2013**, *26*, 10218–10231. [[CrossRef](#)]
32. Anderson, J.L.; Anderson, S.L. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **1999**, *127*, 2741–2758. [[CrossRef](#)]
33. Aksoy, A.; Zhang, F.Q.; Nielsen-Gammon, J.W. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Weather Rev.* **2006**, *134*, 2951–2970. [[CrossRef](#)]
34. Hunt, B.R.; Kostelich, E.J.; Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Phys. D* **2007**, *230*, 112–126. [[CrossRef](#)]
35. Zhang, S. Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: A review. *Clim. Dyn.* **2020**, *54*, 5127–5144. [[CrossRef](#)]
36. Zhang, F.; Snyder, Q.; Sun, J. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **2004**, *132*, 1238–1253. [[CrossRef](#)]

37. Whitaker, J.; Hamill, M. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev.* **2012**, *140*, 3078–3089. [[CrossRef](#)]
38. Wu, X.; Zhang, S.; Liu, Z.; Rosati, A.; Delworth, T. A study of impact of the geographic dependent of observing system on parameter estimation with intermediate coupled model. *Clim. Dyn.* **2013**, *40*, 1789–1798. [[CrossRef](#)]
39. Ruiz, J.J.; Pulido, M.; Miyoshi, T. Estimating model parameters with ensemble-based data assimilation: A review. *J. Meteorol. Soc. Jpn. Ser. II* **2013**, *91*, 79–99. [[CrossRef](#)]
40. Gill, A.E. Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **1980**, *106*, 447–462. [[CrossRef](#)]
41. Zebiak, S.E.; Cane, M.A. A model El Niño-Southern oscillation. *Mon. Weather Rev.* **1987**, *115*, 2262–2278. [[CrossRef](#)]
42. Sluka, T.C.; Penny, S.G.; Kalnay, E.; Miyoshi, T. Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophys. Res. Lett.* **2016**, *43*, 752–759. [[CrossRef](#)]
43. Duc, L.; Saito, K.; Hotta, D. Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 3638–3660. [[CrossRef](#)]
44. Kleeman, R.; Moore, A.M. A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. *J. Atmos. Sci.* **1997**, *54*, 753–767. [[CrossRef](#)]
45. Gaspari, G.; Cohn, S.E. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **1999**, *125*, 723–757. [[CrossRef](#)]
46. Latif, M.; Barnett, T.P.; Cane, M.A.; Flugel, M.; Graham, N.E.; Storch, H.; Xu, J.S.; Zebiak, S.E. A review of ENSO prediction studies. *Clim. Dyn.* **1994**, *9*, 167–179. [[CrossRef](#)]
47. Webster, P.J.; Yang, S. Monsoon and ENSO: Selectively interactive systems. *Q. J. R. Meteorol. Soc.* **1992**, *118*, 877–926. [[CrossRef](#)]
48. Samelson, R.G.; Tziperman, E. Instability of the chaotic ENSO: The growth-phase predictability barrier. *J. Atmos. Sci.* **2001**, *58*, 3613–3625. [[CrossRef](#)]
49. Mu, M.; Duan, W.; Wang, B. Season-dependent dynamics of nonlinear optimal error growth and El Niño-Southern Oscillation predictability in a theoretical model. *J. Geophys. Res.* **2007**, *112*, D10113. [[CrossRef](#)]
50. Jin, E.K.; Kinter, J.L.; Wang, B.; Park, C.K.; Kang, I.S.; Kirtman, B.P.; Kug, J.S.; Kumar, A.; Luo, J.J.; Schemm, J.; et al. Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Clim. Dyn.* **2008**, *31*, 647–664. [[CrossRef](#)]
51. Barnston, A.G.; Tippett, M.K.; L’Heureux, M.L.; Li, S.H.; DeWitt, D.G. Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Am. Meteorol. Soc.* **2012**, *93*, 631–651. [[CrossRef](#)]
52. Lian, T.; Chen, D. An evaluation of rotated EOF analysis and its application to tropical Pacific SST variability. *J. Clim.* **2012**, *25*, 5361–5373. [[CrossRef](#)]
53. Zheng, F.; Fang, X.H.; Zhu, J.; Yu, J.Y.; Li, X.C. Modulation of Bjerknes feedback on the decadal variations in ENSO predictability. *Geophys. Res. Lett.* **2016**, *43*, 12560–12568. [[CrossRef](#)]