

ENSO simulation and prediction in a hybrid coupled model with data assimilation

by

Youmin Tang¹ and William W. Hsieh

Department of Earth and Ocean Sciences

University of British Columbia, Vancouver, B.C. V6T 1Z4, Canada

Revised version for Journal of the Meteorological Society of Japan

September 3, 2002

[keywords: El Niño-Southern Oscillation, data assimilation, neural networks, coupled model]

¹Corresponding author: Youmin Tang, CAOS, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. Email: ytang@cims.nyu.edu

Abstract

With a 3D Var assimilation scheme, several types of observations— sea surface temperatures (SST), sea level height anomalies (SLHA), and the upper ocean 400 meter depth-averaged heat content anomalies (HCA)— were assimilated into a hybrid coupled model of the tropical Pacific. The ocean analyses and prediction skills of the SST anomalies (SSTA) from the assimilation of each type of observation were presented for 1980-1998. SST assimilation, besides improving the simulation of SSTA, also slightly improved the HCA and SLHA simulations in the equatorial Pacific, especially in the east. The ocean analyses with the assimilation of SLHA improved the simulations of SSTA, SLHA and HCA in the equatorial Pacific, while the assimilation of HCA improved the SLHA and HCA simulations.

For ENSO predictions, assimilating SST yielded the best prediction skills for the Niño3 region SSTA at lead times of 3 months or shorter, but severely degraded the predictions at longer lead times. The best Niño3 SSTA predictions for lead times longer than 3 months came from the initializations with the assimilation of HCA and SLHA data. Assimilating SLHA yielded prediction skills for the Niño3 SSTA almost as good as assimilating HCA, indicating considerable potential for improving ENSO predictions from altimetry data.

In this study, a neural network (NN) approach was used to find the nonlinear statistical relations among model variables for the assimilation of HCA and SLHA. Using NN yielded better prediction skills than using multiple linear regression.

1 Introduction

For the ENSO (El Niño-Southern Oscillation) phenomenon (Battisti and Sarachik, 1995; Neelin et al. 1998), prediction skills are dependent on the initialization scheme used in the prediction models. When Zebiak and Cane (1987) first made ENSO predictions with a dynamical coupled model, they used the observed surface wind stress in the tropical Pacific to force the ocean model for several years prior to the prediction. Predictions were then made by running the coupled model forward in time from the established ocean initial conditions. Due to the fast response of the atmosphere to the ocean, the initial conditions of the atmospheric model were a ‘slave’ to the ocean state as expressed through the sea surface temperature (SST). The initialization scheme has been widely adopted in subsequent intermediate coupled models and coupled general circulation models (Balmaseda et al. 1994; Kleeman 1993; Kirtman and Zebiak 1997) and can produce a useful prediction skill of around 0.6 correlation for Niño3 SSTA at the lead time of 6 months.

The above initialization scheme only uses the surface wind stress information. An alternative is to initialize the prediction models taking into account the feedback from the ocean to the atmosphere, or to use other observations, in particular, the direct observations of the ocean state. The hybrid coupled model of Barnett et al. (1993) was initialized with observed SST. Chen et al. (1995) developed an ocean initialization scheme for the Zebiak and Cane (1987) coupled model with the consideration of feedback from ocean to atmosphere through a nudging approach. Furthermore, they assimilated the sea level data using a similar nudging scheme for the coupled model (Chen et al. 1998), improving the forecast skills of the Lamont model. Kleeman et al. (1995) assimilated the 400-m depth-averaged equatorial temperature anomaly into an intermediate complexity coupled model

by using the adjoint variational technique. In their work, the subsurface information was used as proxy data, connected by an empirical relation to the model prognostic variables, for initializing the model states. The results indicated that such an initialization approach can lead to a marked increase in hindcast skills, in particular for long lead times exceeding 8 months. Ji et al. (1998) and Rosati et al. (1997) used the observed three-dimensional temperatures to directly initialize the coupled general circulation models through a 3D Var assimilation system, and found that initializing the ocean by the assimilation of observed subsurface temperature data generally leads to improved prediction skills.

Compared with the field of numerical weather prediction, techniques for data assimilation and initialization in ENSO prediction are still at a relatively early stage of development. A basic issue – which observations are most useful in initializing ENSO predictions, is still not clear for the seasonal prediction community. A major obstacle to answering this question is the expensive hindcast experiments with data assimilation needed by fully coupled models, as a considerable ensemble of experiments seems necessary for a robust comparison. Another difficulty is that oceanic observations are still considered spatially sparse and temporally sporadic, although the ocean observing network for the tropical Pacific was vastly improved during TOGA (Tropical Ocean Global Atmosphere) (McPhaden et al. 1998). The complete and accurate altimetry observations available provide a new opportunity to explore the impact of the sea level data not only in the extra-tropical oceanic variability (Ikeda et al. 1995, Kamachi et al. 2001), but also on ENSO simulation and prediction. The comparisons between the impact of assimilating altimetric sea level heights and assimilating subsurface in situ observations on the initialization for ENSO predictions have recently been addressed by several groups (e.g. Ji et al 2000; Carton et

al. 1996; Segschneider et al. 2000). However, much of these work mainly focused on the ocean analyses and the time periods discussed only cover a few ENSO cycles, limited by the short history of datasets (e.g. the altimetry observations only beginning from 1992). Therefore, a major interest in this paper is to try to assimilate different types of data and examine their impact on the ocean analyses and ENSO prediction skills. A final objective is to show that data assimilation with the use of nonlinear neural networks can yield higher ENSO prediction skills when compared to the use of linear regression.

This paper is structured as follows: Section 2 briefly describes the hybrid coupled model and the assimilation scheme. Section 3 introduces the data and proposes strategies for assimilating SLHA and HCA into the ocean model. The ocean analyses from assimilating different types of data and the impact of initializations from the ocean analyses on ENSO predictions are presented in Sections 4 and 5, respectively, followed by summary and discussion in section 6.

2 The hybrid coupled model and the assimilation system

2.1 The hybrid coupled model

The hybrid-coupled model (HCM) combines a dynamical ocean model with a statistical atmospheric model for coupled interaction studies (Barnett et al. 1993; Syu and Neelin 2000; Balmaseda et al. 1994). Although fully dynamical coupled ocean-atmosphere models have been well developed and more widely used in recent years, HCM still has some merits, such as (1) the climate drift problem is avoided; (2) lower computing costs than

a full CGCM (coupled general circulation model) which allow us to run a considerable ensemble of hindcast experiments; (3) comparable or even better ENSO prediction skills relative to a CGCM (Palmer and Anderson 1994; Kang and Kug 2000; Barnston et al. 1999).

The hybrid coupled model used in this study is composed of a nonlinear dynamical ocean model and a nonlinear empirical atmospheric model (built from neural networks), and identical to the NHCM (Nonlinear Hybrid Coupled Model) in Tang (2001). The ocean model is one of intermediate complexity, originated from Anderson and McCreary (1985) and Balmaseda et al. (1994, 1995), but extended to six active layers in this study. The $1.5^\circ \times 1.5^\circ$ resolution model, covering the area $30^\circ\text{N} - 30^\circ\text{S}$ and $123^\circ\text{E} - 69^\circ\text{W}$, consists of the depth averaged primitive equations in six active layers (with initial thickness of 100 m, 175 m, 250 m, 320 m, 400 m and 500 m from top to bottom), overlying a deep inert layer. The model allows for an exchange of mass, momentum and heat at each interface by vertical mixing parameterized as entrainment. The time step for integration is two hours. The detailed formulation of the ocean model is presented in the Appendix.

A control experiment is first made with the uncoupled ocean model, forced by the Florida State University (FSU) observed wind stress (Goldenberg and O'Brien 1981) from 1961-1989. The upper ocean heat content (HC) from the control run is then used as predictors in a nonlinear empirical model for reconstructing the wind stress at zero lead time. The first 3 PCs (Principal Components, i.e. the EOF time coefficients) for the HC were chosen as predictors and the leading three wind stress (zonal or meridional) PCs as predictands. A separate empirical model was used to predict each wind stress PC. Feed-forward neural networks (NN) (Hsieh and Tang, 1998) were used to construct the

nonlinear atmospheric models.

To alleviate artificial skills, the cross-validated scheme (von Storch and Zwiers 1999) was used to develop the atmospheric models during 1964-1989. A cross-validation procedure involves dividing the data record into several segments, selecting one segment for the test data and the rest for training data. The NN model was built using the training data only, and model predictions were tested on the test data. Next, another segment was selected as the test data, and a new model version was built. This was repeated until the entire data record had been used for testing. The cross-validation scheme ensures that no training data are used for testing the prediction skills, hence the artificial skill associated with the overfitting problem can be effectively eliminated. Further details of the ocean model, the atmospheric model and the coupled model can be found in Tang et al (2001).

The model heat content is defined here using the same formula as Balmaseda et al. (1994, 1995):

$$HC_i = \frac{h_i(T_i - T_7)}{\sum_{i=1}^6 H_{\text{init}}(i)}, \quad HC = \sum_{i=1}^2 HC_i, \quad (1)$$

where T_7 is the temperature of the bottom layer, and $H_{\text{init}}(i)$ is the initial thickness of layer i . Averaged over 30 years of simulation, the climatological monthly model depths in the top two layers range from 89m to 144m in the first layer, and from 134m to 188m in the second layer, with variations in season and in space.

The atmospheric and ocean models are coupled as follows: The anomaly pattern of HC (with respect to the monthly climatology) from the ocean model is projected onto the first three EOF spatial modes to extract the predictors. The atmospheric model (NN) is then used to reconstruct the wind stress anomalies. The reconstructed wind stress anomalies are added to the climatological wind stress field to force the ocean model. The coupling is

done once every 15 days (the coupled model is not very sensitive to the coupling interval varying between one day to one month).

2.2 The assimilation system

The data assimilation system is the 3D Var assimilation system described by Derber and Rosati (1989). Since the vertical correlation is ignored in the observational error covariance matrix as in Derber and Rosati (1989), it is actually a 2D Var assimilation system, though we will still use the common name 3D Var in this paper. The 3D Var assimilation scheme has been used in operational ENSO forecasts at NCEP (National Center for Environmental Predictions) and at GFDL (Geophysical Fluid Dynamics Laboratory) (Ji et al. 1995, 2000; Rosati et al. 1997). The 3D Var minimizes the cost function

$$I = \frac{1}{2} \mathbf{T}^T \mathbf{E}^{-1} \mathbf{T} + \frac{1}{2} (\mathbf{D}(\mathbf{T}) - \mathbf{T}_0)^T \mathbf{F}^{-1} (\mathbf{D}(\mathbf{T}) - \mathbf{T}_0) \quad (2)$$

The superscript T represents the transpose, \mathbf{T} is an N component vector containing the correction to the guess field (the guess field being generated by the model before assimilating the latest data), \mathbf{E} is an estimate of the $N \times N$ first guess error covariance matrix, \mathbf{D} is a simple bilinear operator interpolating from the model grid to the observational stations, \mathbf{T}_0 is an M component vector containing the difference between the observations and the interpolated guess field, and \mathbf{F} is an estimate of the $M \times M$ observational error covariance matrix, where N and M denote the number of model grid points and the number of the observational stations respectively. The first term of the functional ($\mathbf{T}^T \mathbf{E}^{-1} \mathbf{T}$) is a measure of the fit of the corrected field to the guess field, while the second term measures the fit of the corrected field to the observations. The result is

a weighted average of the guess field (which contains information from an earlier period) and the observations.

The estimated spatial error covariance matrices, \mathbf{E} and \mathbf{F} determine the spatial structure and amplitude of the correlation field. A simple Gaussian function with the homogeneous and isotropic assumptions is used to determine \mathbf{E} while ignoring the vertical correlations as in Derber and Rosati (1989), i.e. the horizontal covariance between any two points is given by

$$ae^{-r^2/(b^2\cos\phi)}, \tag{3}$$

where ϕ is the latitude of the grid point, r the distance between any two points, $a=0.01$, and $b = 570$ km, as in Derber and Rosati (1989).

For the observational error covariance matrix \mathbf{F} , we assume that the observational errors are not correlated, i.e. the matrix \mathbf{F} has only diagonal elements which are the observational error variances. For SST and SLHA, each element of the diagonal error covariance matrix is constant, with the value $(0.5^{\circ}\text{C})^2$ for SST. For the HC anomaly, the error variances assigned are $(0.25^{\circ}\text{C})^2$ in the western Pacific and $(0.5^{\circ}\text{C})^2$ in the eastern Pacific due to less observational stations in the east. These values are obtained empirically from sensitivity experiments, and following recommendations from previous studies (e.g. Ji et al, 1995). The observational error variances used in the SLHA assimilation will be given in section 3.3 as a different approach is used for SLHA assimilation.

The function I in (2) is minimized using a preconditioned conjugate gradient algorithm (Gill et al. 1981). The preconditioning in the algorithm is supplied by the \mathbf{E} matrix, which allows the solution to be found without directly inverting the \mathbf{E} matrix. Further details about the algorithm can be found in Derber and Rosati (1989) and Gill et al. (1981).

The assimilation system also includes data quality control. The initial gross check is to ensure that the data used are not over land. This is implemented by one land-ocean mask in which the grid weights are set to 1 over ocean and 0 over land. Before the guess field and observations enter the assimilation system, they were both mask-filtered. When the magnitude difference between the guess field and the observations is greater than 25% of the magnitude of the guess field, the observations are eliminated. This might lose effective corrections for poor simulations in a few model grid points, but ensures the ocean model is gradually adjusted, without problems arising from the imbalance between the velocity and density fields. The data are inserted daily into the model with a smooth transition at the end of each month.

3 Methodology for assimilating different types of data

3.1 Data

The data used in this study are the monthly mean SST obtained from COADS (Comprehensive Ocean-Atmospheric Data Set, Smith et al. 1996) with 2° lat. by 2° lon. resolution, and the monthly 2° by 2° mean surface wind stress from FSU. For SLHA data, we use the NCEP reanalysis data set of 1.5° lat. by 1° lon. instead of the observations, because it has a longer record and higher spatial resolution than the short altimetry observations or the sparse tide gauge data. This reanalysis data set has recently been used for ENSO predictions (e.g. Syu and Neelin 2000; Kang and Kug 2000). The 20° isotherm depth at the equator used in next section is also from the reanalysis data set. For brevity, we call the SLHA from the reanalysis data set the NCEP SLHA.

The monthly 400-m depth-averaged heat content anomalies (HCA) are from the data set of subsurface temperature and heat content provided by the Joint Environmental Data Analysis Center at the Scripps Institution of Oceanography. This data set consists of all available XBT, CTD, MBT and hydrographic observations, optimally interpolated by White (1995) to a three-dimensional grid of 2° lat. by 5° lon., and 11 standard depth levels between the surface and 400m.

SST can be directly assimilated into the model because their model climatology is similar to their observed climatology and because they are also prognostic variables in the model. With SLHA and HCA, the situation is more complicated.

3.2 HCA observations

An important component of the coupled ocean-atmosphere system on the interannual timescale is the upper ocean heat content, which is the source of memory for the system. Many studies have shown (e.g. Zebiak and Cane 1987; Rosati et al. 1997) that the fluctuations in upper-ocean heat content are both systematic and significant in the evolution of ENSO.

According to Eq. (1), the model HCA can be computed at each model step; however, assimilating the HCA data is not easy, as HCA is not a prognostic model variable. When Kleeman et al. (1995) and Moore and Anderson (1989) assimilated subsurface temperature into their ocean models, they used a linear empirical relation to transfer the corrections in the subsurface thermal fields to the corrections of the model prognostic variables. Here we will use a similar method to project the HCA analyses to four model prognostic variables, — SSTA, the first layer thickness anomalies (H1A), the second layer temperature anoma-

lies (T2A) and the second layer thickness anomalies (H2A). Instead of linear regression, we use nonlinear regression by NN to find the empirical relations between HCA and these four model variables, as there is no a priori reason to believe these relationships to be linear.

The procedure to model these relations is very similar to the construction of atmospheric models in Section 2: (1) An EOF analysis is applied to the control model HCA and to each of the four prognostic variables during the period 1964-1990. The first three modes are retained to construct the empirical relation, as they explained over 60% of the variance for all variables. (2) An ensemble of neural networks (25 NNs used here) is trained where the inputs to each NN are the 3 PCs of the HCA and the output is one of the PCs of a prognostic variable. Three hidden neurons are used in each NN. The final relation is the ensemble average over 25 NNs. The cross-validated scheme is used to obtain these empirical equations as done in the construction of the atmospheric model. A total of 12 ensembles of NNs are needed to reconstruct the 3 leading PCs of the 4 prognostic variables from HCA. (3) As the observed heat content is the integral of the temperatures over the upper 400 meters, while the model HC is defined as the integral of the temperatures over the first two layers (275 meters), the observed HCA values are scaled to match the standard deviation of the model HCA before being assimilated.

Next, we perform a consistency check: HCA from the control experiment was performed to estimate the prognostic variables, which were then used to recompute HCA by Eq. (1). Fig. 1 shows a time-longitude plot of the HCA along the equator during 1980-1998, taken from observations, control experiment and recomputed from the prognostic variables SSTA, H1A, T2A, H2A. As shown in Fig. 1c, the recomputed HCA agreed well

with the original control experiment values, and both agreed fairly well with the observed values.

3.3 SLHA observations

As sea level height is usually a diagnostic variable in intermediate complexity ocean models, the SLHA observation cannot be directly assimilated into the ocean models. Even when sea level height is a prognostic variable in models, it is not sensible to correct the sea level height directly. So a similar approach as before, i.e. using linear regression to transfer SLHA corrections to the corrections in the other model prognostic variables, is often used for SLHA assimilation (e.g. Ji et al. 1998, 2000; Carton et al. 1996). We first performed the SLHA assimilation using the same approach as in our HCA assimilation. Unfortunately, this approach did not work well in assimilating SLHA into our ocean model, with the prediction skills of Niño3 SSTA even degrading at all lead times (not shown). The reason is not well understood, but the systematic errors of the SLHA amplitude simulation by the low vertical resolution model is probably one major cause. The magnitude of the systematic errors vary spatially in the tropical Pacific, and one cannot remove the systematic errors by uniformly scaling the simulated SLHA. An alternative approach is needed.

Since large-scale variations in sea level height are in effect determined by dynamical and thermodynamical processes of the upper ocean, there are good relations among SLHA, the displacement of thermocline and SSTA as shown in Fig. 2. Also, the top thickness anomalies H1A of the model are mainly dominated by the lowest-order baroclinic mode and fairly well reflect the thermocline perturbations as shown in Fig. 3. The alternative

approach is to derive proxy data of the SSTA and H1A from the NCEP SLHA.

An EOF analysis is first performed separately for the control experiment SLHA, SSTA and H1A. The modeled SLHA are obtained by diagnosing the modeled sea surface pressure (Tang 2001). The statistical relations of SLHA to H1A and to SSTA are extracted with a nonlinear NN approach using the first three PCs for each variable — the first 3 PCs accounted for a total of 65% of the variance for all three variables.

Then from the leading 3 NCEP SLHA PCs, the NNs gave the 3 leading PCs for the SSTA or the H1A, which when combined with the spatial EOF patterns yielded the proxy SSTA and H1A fields. In terms of agreement with the observations, the proxy SSTA is much better than the SSTA from the control experiment (Fig. 4a), justifying that the proxy SSTA can be used to correct the model SSTA. There are observations to compare directly with the proxy H1A; however, Fig. 4b shows that the proxy H1A indeed better simulates observed ENSO signals than the H1A from the control experiment. For example, the control experiment simulates the peak time of 1997 El Niño around early 1997, too early compared with other observed variables (e.g. SSTA).

With this approach, the assimilation of SLHA is transferred to the assimilation of proxy SSTA and proxy H1A. The reason that the proxy SSTA data is used instead of the observed SSTA in this approach is that we want to only use SLHA information, and to investigate the impact of assimilating SLHA on ENSO simulations and predictions. The observational error covariance for proxy SSTA and H1A was set to $(0.5^\circ\text{C})^2$ and $(1\text{m})^2$.

3.4 Assimilation approaches

The assimilation is performed only for the ocean model. As the SLHA data starts from 1980 only and the HCA data has relatively poor quality prior to 1980, all experiments are set to start from 1980. As in the control experiment, the ocean model is forced with the observed FSU wind stress, but monthly oceanic data are assimilated during January 1980 to April 1998. Predictions are then made by running the coupled model forward in time from the established ocean initial conditions for 15 months. The assimilation domain is confined to the equatorial Pacific, $15^{\circ}\text{S} - 15^{\circ}\text{N}$, for all experiments.

4 Ocean analyses

First, we will examine the impact of data assimilation on ENSO simulation through comparing the ocean analyses from the assimilation of different data sets. Three physical variables, SSTA, SLHA and HCA, are examined in the ocean analyses. As the ocean analyses are also the initial fields of the predictions to be discussed in the next section, the following analyses will actually explore the features of the initial fields.

The correlations and RMSE (root mean square error) between the observed SSTA and the modeled SSTA from the control experiment and the oceanic analyses are shown in Fig. 5. A common characteristics in all experiments except SST assimilation is that good simulated skills only appear in the eastern-central equatorial Pacific. This is probably because the assimilation cannot effectively correct the parameterization of heat flux which mainly determines the SST variations in the western Pacific. The poor simulated skills of SSTA in the western Pacific are common defects in almost all ocean models including

OGCMs (Barnett et al. 1993; Chen et al. 1997). As in Fig. 5, the oceanic analyses with the assimilation of SST and SLHA are significantly better than the control experiment in simulating the tropical Pacific SSTA. The most pronounced improvement in the tropical Pacific SSTA simulation is from SST assimilation, with high correlation (over 0.8) between observations and simulations for almost the whole assimilation domain. Since the model simulations are compared to the same data that were directly assimilated, it is not surprising that such a high correlation was achieved. The analyses with HCA assimilation brought little improvement to the SSTA simulation, even slightly worse in the equatorial eastern boundary.

Fig. 6 shows the correlations and RMSE between the NCEP SLHA and the modeled SLHA from the control experiment and the oceanic analyses. SLHA simulations in all experiments share a similar feature, i.e. good correlation of over 0.6 cover the whole equatorial belt in contrast with the good correlation found only in the equatorial eastern Pacific for SSTA simulations. In terms of correlation, the improvements in the SLHA simulation are mainly in the central equatorial Pacific with SLHA assimilation, and in the eastern equatorial Pacific with SST and HCA assimilation. The assimilation of SLHA, however, generates a bigger RMSE in the eastern equatorial Pacific compared with the control experiment. This is probably associated with the use of proxy data, since the statistical models for producing the proxy data generally reduce the predicted variance.

The comparisons between observations and analyses in terms of HCA simulation are shown in Fig. 7. The best correlation coefficients appear in the analysis with HCA assimilation as expected. As we have assimilated HCA into the ocean model through an indirect approach, the good improvement of HCA simulation from this ocean analysis

also provides some justification for this approach. The analyses from the assimilation of SLHA or SST also generate improvement for the HCA simulation in the equatorial eastern Pacific.

Fig. 8 shows the evolution of the observed and modeled SSTA in the Niño3 region in the eastern equatorial Pacific. While all experiments captured the major observed ENSO signals, the HCA assimilation has the better capability for generating the anomalous extremes than SLHA assimilation and the control experiment, although it sometimes exaggerates the extrema simulations. The common deficiencies for all experiments (except that with SST assimilation) are the delay in simulating the occurrence of the 1982-83 warm event, and the relatively weak simulated amplitude for the 1997 event. However, the phase delay in 1982 was much alleviated with HCA assimilation. In addition, the much colder simulation than observation in the control experiment for 1995-1997 were also alleviated in the analyses with the assimilation of HCA and, to lesser extents, SLHA. The evolution of the observed and modeled Niño3 SLHA and Niño3 HCA (not shown) also manifest similar features as the Niño3 SSTA.

5 ENSO prediction

Next we examine the predictions by the NHCM. A total of 73 prediction cases, initialized from April 1980 to April 1998, were run starting at three months intervals (1 January, 1 April, 1 July, 1 October), and continued for 15 months.

Fig. 9a shows the NHCM predictive skills initialized from the control experiment and the ocean analyses with the assimilation of several data sets, where the predicted Niño3 SSTA is compared with the observed values. All predictions beat persistence (not shown)

at lead times greater than 2-3 months. Generally, all suffered an initial shock in the first 2-3 months, leading to lower skills than persistence. The correlation skills of the control experiment decline with lead time, reaching a minimum at 9 months, beyond which the skills rebound and stabilize till a lead time of 15 months. The best overall predictive skills were attained by initializations with the assimilation of HCA, with correlation skills of above 0.7 at a lead time of 6 months, and above 0.6 for all lead times. Not far behind were the second best results attained by the assimilation of SLHA. Finally, SST assimilation, which was the best performer at 1-2 month lead times, did much worse than the control experiment at lead times beyond 9 months.

Contrasted with the considerable improvement of the correlation skills in the predictions initialized with SLHA and HCA assimilation, the improvement in the RMSE from these initializations is less impressive (Fig. 9b).

Overall, the predictions initialized with HCA and SLHA assimilation are better than the predictions initialized from the control experiment. The SST assimilation attains very good prediction skills for short lead times, but underperforms the control for lead times beyond 6 months.

Fig. 10 shows the predicted Niño3 SSTA initialized from the control experiment and the different ocean analyses at the lead time of 6 months and 12 months. It was found that the initializations from the ocean analyses with the assimilation of SST, SLHA and HCA— especially HCA and SLHA— seem to give better predictions for ENSO warm events. But it seems that the initializations from these ocean analyses can more easily produce spurious ENSO cold events, especially for the predictions initialized with SST assimilation.

We next show several prediction cases for two typical ENSO events, the 1982/83 and 1997 El Niño. The predictions started from the January of the event year to the following January.

The predictions of the Niño3 SSTA initialized from various assimilation experiments are plotted in Fig. 11 for the two events. Compared with the observed SSTA, all experiments successfully predicted the two anomalous warmings, except for the SST assimilation in the 1982/83 event. However, except for the experiments with SLHA and HCA assimilation, the other predictions suffered a phase delay in the peak time, especially for the predictions initialized from the control experiment. Once again, the initializations with the SLHA and HCA assimilation generated the best predictions in terms of the phase and amplitude.

A key step in the assimilation of HCA and SLHA is the construction of statistical relations either for transferring assimilated information to model prognostic variables or for generating proxy data. Here NN is used instead of traditional linear regression (LR). To explore whether NN has any advantage over LR, the two NHCM assimilation experiments for HCA and SLHA were repeated using LR.

As seen in Fig. 12, the prediction skills are generally sensitive to the statistical relation used in the assimilation, in particular for HCA assimilation. Because the ocean analyses were derived from the statistical relations in HCA assimilation, in contrast to SLHA assimilation where the ocean analyses are the weighted average of model simulations and observation, the HCA assimilation could have a larger initial shock than the SLHA assimilation. The initial shock might be more affected by the assimilation scheme itself than the choice of linear or nonlinear statistical relation, so that the differences between

NN and LR could be subtle in the first few months of prediction for HCA assimilation as shown in Fig. 12 (left panel). After the initial shock has dissipated after a few months, the advantages of NN over LR are then manifested. In contrast, the initial shock is small for SLHA assimilation, so the advantages of NN could be seen at short lead times (Fig. 12, right panel). At longer lead times, NN did not show advantages relative to LR in SLHA assimilation.

The more significant differences between NN and LR in HCA assimilation than in SLHA assimilation is probably associated with their methodology. In SLHA assimilation, only SSTA and H1A variables are related to SLHA, whereas in HCA assimilation, additional deeper variables (T2A and H2A) which are less linearly related to the surface conditions, are related to HCA. Hence, stronger nonlinear NN relations were likely used in HCA assimilation than in SLHA assimilation, thereby accentuating the difference between NN and LR.

The finite sample size implies some uncertainty in the computed correlation coefficient. Error bars in Fig. 12 were obtained by a bootstrap method (Yuval, 2001). The differences in the correlation skills appear to result from the assimilation methodology than from the uncertainty of the finite sample size.

6 Summary and discussion

Using a 3D Var assimilation scheme, we have examined the effects from assimilating different types of data on the ocean analyses and on ENSO prediction skills. The data assimilated were the SST, SLHA and the upper ocean 400 meter depth-averaged HCA, during 1980-1998. Neural network (NN) techniques have been used to relate data for

diagnostic variables (e.g. HCA and SLHA) to prognostic variables in the model.

Three variables, SSTA, SLHA and HCA, were chosen to validate the ocean analyses. The analyses with SST assimilation generated the best performance for SSTA simulation in the whole assimilation domain, and some improvements for SLHA and HCA simulation in the eastern and central equatorial Pacific. The analyses with the assimilation of SLHA improved the simulations of all three variables in the eastern and central equatorial Pacific. The assimilation of HCA improved the ocean analysis mainly for HCA itself, and slightly for SLHA in the eastern and central equatorial Pacific.

For the Niño3 region SSTA predictions, HCA and SLHA assimilation were the best in improving the prediction skills—the improvement from SST assimilation was the best only for lead times of 3 months or shorter; at longer lead times, SST assimilation degraded the predictions.

One interesting result is that the SST assimilation can lead to relatively good ocean analysis but poor prediction skills at long lead times. There are several possible reasons why the assimilation of SST observations cannot benefit ENSO predictions at long lead times. The first is that the methodology cannot effectively correct the subsurface thermodynamical structure. The thermocline is mainly affected by the atmospheric wind stress and the direct impact of SST on it is of little significance. Another reason is that there exist large systematic differences in the spatial distribution of variances between the model SST field and the observed SST field, which can be seen in their EOF modes (not shown). Compared with the observed SST modes, the modeled SST modes appear to be more narrowly confined to the equator, with less variability near the eastern boundary. These are common defects found in many ocean models including OGCMs. With the

assimilation of the observed SST, the structure of the model SST is quickly forced to resemble its observational counterpart; however, the model adjustment is relatively slow, especially for the adjustment of the thermocline, which mainly determines the variability of SST anomalies in the equatorial central and eastern Pacific, resulting in imbalances between the dynamical and thermal fields, i.e., the observational forcing to the model SST is far faster than the adjustment of dynamical fields. A balanced state of the model at the initial time is very important for good prediction skills at longer lead times. This could be suggested by Fig. 5, where the SST corrections are almost ‘perfect’, and Figs. 6 and 7, where the improvements in SLHA and HCA are far less perfect. One cannot expect to effectively correct model biases by data assimilation in most situations, especially by three dimensional assimilation without posing any constraint conditions. Corrections to systematic biases should mainly rely on the improvement of model physics.

While this work has shed some light on the effects of assimilating various types of data on ENSO simulation and prediction, many questions remain. For instance, would assimilating more than one type of data be able to provide even better initial conditions for ENSO prediction? How are the results shown in this paper sensitive to factors other than the assimilated data— e.g. the assimilation scheme, the assimilation parameters, the model and observational error covariance matrices, and the methods for validation? The importance of these factors might even exceed the type of assimilated data in determining the performance of the assimilation experiments. We have explored some of these factors by running some sensitive experiments— in a set of sensitivity experiments, the one generating the best Niño3 SSTA predictions was selected as the appropriate one. Sensitivity experiments have been performed for choosing the observational error covariance matrix,

and the universal values a and b in Eq. (3).

That the methodology of assimilation could have considerable impact is revealed in the case of HCA and SLHA assimilation. The neural network tended to yield higher forecast skills than linear regression, implying that the nonlinearity of ocean variables cannot be ignored in the assimilation. In addition, the approach used successfully in HCA assimilation did not work well for SLHA assimilation, so a rather different procedure was used for SLHA assimilation in this study. This also suggests that the comparisons with the assimilation of SST may not be entirely fair, as it was directly assimilated into the model, unlike the assimilation of SLHA and HCA, where statistical relationships from NNs were used. Nevertheless this exploratory work has (a) allowed us to implement data assimilation schemes that are likely to improve the ENSO prediction skills of our hybrid coupled model, which has been used to issue routine operational ENSO forecasts, and (b) provided two applicable nonlinear methods for assimilating HCA and SLHA respectively into intermediate complexity equatorial coupled models.

7 Appendix: Ocean Dynamical Model

Like the two-layer model of Balmaseda et al (1994), this 6-layer model includes explicit nonlinear thermodynamics for all layers, accounting for horizontal advection, vertical heat transport, diffusion and, in the first layer, surface heat flux. Besides the physical processes considered in the two-layer model, the 6-layer model includes a viscosity between these layers to account for wind driven Ekman effects, simple connective mixing between 2 adjacent layers due to upwelling and horizontal advection.

The model equations are as follows;

$$\left(\frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x} + v_i \frac{\partial}{\partial y} + f \mathbf{k} \wedge \right) h_i \mathbf{u}_i = -\frac{h_i}{\rho_0} \nabla p_i + \mathbf{M}_i + \frac{\tau}{\rho_0} + \nu \nabla^2 (h_i \mathbf{u}_i) + \mathbf{V}_i, \quad (4)$$

$$\left(\frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x} + v_i \frac{\partial}{\partial y} \right) T_i = Q_s + H_i + \kappa \nabla^2 T_i, \quad (5)$$

where layer subscript $i = 1, \dots, N$. The symbols u and v are the zonal and meridional components of velocity respectively, p is the pressure, h is the layer thickness, bf τ is the wind stress, T is temperature, f is the Coriolis parameter given by $f = \beta y$, ρ_0 is a constant density, and ν , d and κ are coefficients of momentum, thickness and heat diffusivity. The terms D_i , \mathbf{M}_i , H_i are mass, momentum and heat exchanges between layers and are discussed below. The term Q_s is the surface heat flux determined as for the 2-layer model and \mathbf{V}_i is a vertical viscosity used if the Ekman layer is resolved. \mathbf{V}_i is parameterized as

$$\mathbf{V}_i = A_i (\mathbf{u}_{i-1} - \mathbf{u}_i) + A_{i+1} (\mathbf{u}_{i+1} - \mathbf{u}_i), \quad \text{where } A_{i+1} \leq A_i$$

The pressure gradients are

$$\frac{1}{\rho_0} \nabla p_i = g \alpha \nabla \left(\sum_{j=1}^{i-1} h_j (T_i - T_{N+1}) + \sum_{j=i}^N h_j (T_j - T_{N+1}) \right) - g \alpha \left(\frac{h_i}{2} + \sum_{j=1}^{i-1} h_j \right) \nabla T_i \quad (6)$$

where g and α are the gravitational constant and the thermal expansion coefficient respectively. The pressure at sea surface is

$$p_s = -g \rho_{N+1} \left(H_0 - \sum_{j=1}^N h_j \right) - g \sum_{j=1}^N \rho_j h_j, \quad (7)$$

and H_0 is the total depth of the model (a constant), and ρ_j is the density of the j th layer.

The upwelling, ϵ_{ei} , and downwelling, ϵ_{di} , velocities between layers i and $i + 1$ are parameterised as follows;

$$\epsilon_{ei} = (H_i - h_i)^2 / t_{ei} H_i, \quad \text{if } h_i \leq H_i \quad \text{else } \epsilon_{ei} = 0, \quad (8)$$

$$\epsilon_{di} = -(H_i - h_i)^2/t_{di}H_i, \quad \text{if } h_i \geq H_i \quad \text{else} \quad \epsilon_{di} = 0,$$

where H_i is the nominal layer depth towards which entrainment or detrainment tend to force layer i , and t_{ei} and t_{di} are the entrainment and detrainment time scales respectively.

The total vertical velocity may therefore be expressed as

$$\epsilon'_i = \frac{1}{2}(\epsilon_{ei} + |\epsilon_{ei}|) + \frac{1}{2}(\epsilon_{di} - |\epsilon_{di}|).$$

The thickness of the i th layer can be changed by mass exchanges at both its top and bottom surfaces, with layer $i - 1$ above and $i + 1$ below. The total mass exchanged in (6) is given by

$$D_i = \epsilon'_i - \epsilon'_{i-1}.$$

Note that within each layer the parameters only control entrainment or detrainment with the layer below. Any entrainment or detrainment with the layer above is determined by the thickness of the layer above.

The momentum exchanges for layer i are

$$\mathbf{M}_i = \epsilon'_i \mathbf{u}_{i+1} - \epsilon'_{i-1} \mathbf{u}_i \quad \text{for } \epsilon'_i \geq 0, \epsilon'_{i-1} \geq 0, \quad (9)$$

$$\mathbf{M}_i = \epsilon'_i \mathbf{u}_{i+1} - \epsilon'_{i-1} \mathbf{u}_{i-1} \quad \text{for } \epsilon'_i \geq 0, \epsilon'_{i-1} \leq 0,$$

$$\mathbf{M}_i = \epsilon'_i \mathbf{u}_i - \epsilon'_{i-1} \mathbf{u}_{i-1} \quad \text{for } \epsilon'_i \leq 0, \epsilon'_{i-1} \leq 0,$$

$$\mathbf{M}_i = \epsilon'_i \mathbf{u}_i - \epsilon'_{i-1} \mathbf{u}_i \quad \text{for } \epsilon'_i \leq 0, \epsilon'_{i-1} \geq 0.$$

The heat transport term related to upwelling and downwelling is

$$H_i = -\epsilon'_i \lambda_i (T_i - T_{i+1})/h_i - \epsilon'_{i-1} \lambda_{i-1} (T_{i-1} - T_i)/h_i \quad \text{for } \epsilon'_i \geq 0, \epsilon'_{i-1} \leq 0, \quad (10)$$

$$H_i = 0 \quad \text{for} \quad \epsilon'_i \leq 0, \epsilon'_{i-1} \geq 0,$$

where λ_i is a parameter controlling the effective vertical temperature gradient. To make total heat content conservative, λ_i is taken to be 1.

Acknowledgements

Y. Tang would like to thank Xu Li and Keith Haines for helpful discussions, and K. Haines, D.L.T. Anderson and M.A. Balmaseda for providing the ocean model used in the hybrid coupled model. This work was supported by research and strategic grants to W. Hsieh from the Natural Sciences and Engineering Research Council of Canada.

References

- Anderson D. L. T., and J. P. McCreary, 1985: Slowly propagating disturbances in a coupled ocean-atmosphere model. *J. Atmos. Sci.*, **42**, 615-629.
- Balmaseda, M. A., D. L. T. Anderson, and M. K. Davey, 1994: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres, *Tellus*, **46(A)** (4), 497-511.
- , M. K. Davey, and D. L. T. Anderson, 1995: Decadal and seasonal dependence of ENSO prediction skill. *J. Climate*, **8**, 2705-2715.
- Barnett, T. P., M. Latif, N. E. Graham, M. Flügel, S. Pazan and W. White, 1993: ENSO and ENSO related predictability, Part1: Prediction of equatorial sea surface temperature with a hybrid coupled ocean-atmosphere model. *J. Climate*, **6**, 1545-1566.
- Barnston, A. G., M. H. Glantz, and Y. X. He, 1999: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997-98 El Ni episode and the 1998 La Nina onset. *Bull. Amer. Meteor. Soc.*, **80**, 217-243.
- Battisti, D.S., and E.S. Sarachik, 1995: Understnading and predicting ENSO. *Rev. Geophys.*, **33**, 1367-1376.
- Carton, J. A., B. S. Giese, X. Cao and L. Miller, 1996: Impact of altimeter, thermistor, and expendable bathythermograph data on retrospective analyses of the tropical Pacific ocean. *J. Geophys. Res.*, **101**, 14147-14159.

- Chen, D., S. E. Zebiak, A. J. Busalacchi and M. A. Cane, 1995: An improved procedure for El Niño forecasting: Implications for predictability. *Science*, **269**, 1699-1702.
- , — and M. A. Cane, 1997: Initialization and predictability of a coupled ENSO forecast model. *Mon. Wea. Rev.*, **125**, 773-788.
- , M. A. Cane, E. Zebiak and A. Kaplan, 1998: The impact of sea level data assimilation on the Lamont model prediction of the 1997/98 El Niño. *Geophys. Res. Letter*, **25**, 2837-2840.
- Derber, J., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333-1347.
- Gill, P. E., W. Murray and M. H. Wright, 1981: *Practical Optimization*. Academic Press, 401pp.
- Goldenberg, B. and J. J. O'Brien, 1981: Time and space variability of tropical Pacific wind stress, *Mon. Wea. Rev.*, **109**, 1190-1207.
- Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.*, **79**, 1855-1870.
- Ikeda, M., G. Evensen, and L. Cong, 1995: Comparison of sequential updating, Kalman filter and variational-methods for assimilating Rossby waves in the simulated GEOSAT altimeter data into a quasi-geostrophic model. *J. Marine Systems*, **6**, 15-30.
- Ji, M., A. Leetmaa and J. Derber, 1995: An ocean analysis for seasonal to interannual climate studies. *Mon. Wea. Rev.*, **123**, 460-481.

- , D. W. Behringer, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization –PART II- The coupled model. *Mon. Wea. Rev.*, **126**, 1022-1034.
- , R. W. Reynolds, and D. W. Behringer, 2000: Use of TOPEX/Poseidon sea level data for ocean analyses and ENSO prediction: some early results. *J. Climate*, **13**, 216-231.
- Kamachi, M., T. Kuragano, N. Yoshioka, J. Zhu, and F. Uboldi, 2001: Assimilation of satellite altimetry into a western North Pacific operational model. *Adv. Atmos. Sci.*, **18**, 767-786.
- Kang, I., and J. Kug, 2000: An El Niño prediction system using an intermediate ocean and a statistical atmosphere. *Geophys. Res. Letter*, **27**, 1167-1170.
- Kirtman, B., and S. E. Zebiak, 1997: ENSO simulation and prediction with a hybrid coupled model. *Mon. Weather Rev.*, **125**, 2620-2641.
- Kleeman, R., 1993: On the dependence of hindcast skill in a coupled ocean-atmosphere model on ocean thermodynamics. *J. Climate*, **6**, 2012-2033.
- , A. Moore, and N. R. Smith, 1995: Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean-atmosphere forecast model. *Mon. Wea. Rev.*, **123**, 3103-3113.
- McPhaden, M. J., and Coauthors, 1998: The Tropical Ocean Global Atmosphere (TOGA) observing system: A decade of progress. *J. Geophys. Res.*, **103** (C7), 14 169-14 240.

- Moore, A. M., and D. L. T. Anderson, 1989: The assimilation of XBT data into a layer model of the tropical Pacific ocean. *Dynamics of Atmospheres and Oceans*, **13**, 441-464.
- Neelin, J.D., D.S. Battisti, A.C. Hirst, F.F. Jin, Y. Wakata, T. Yamagata, and S.E. Zebiak, 1998: ENSO theory. *J. Geophys. Res.*, **103**, 14261-14290.
- Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting -A review paper, *Q. J. Roy. Met. Soc.*, **7**, 755-793.
- Perigaud, C., Cassou C., Dewitte B., Fu L-L and Neelin D. 2000: Using Data and Intermediate coupled models for seasonal-to-interannual forecasts. *Mon. Wea. Rev.*, **128**, 3025-3049.
- Rosati, A., K. Miyakoda and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.*, **125**, 754-772.
- Segsneider, J., D. L. T. Anderson, and T. N. Stockdale, 2000: Towards the use of altimetry for operational seasonal forecasting. *J. Climate*, **13**, 3115-3138.
- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403-1420.
- Syu, H-H and D. Neelin, 2000: ENSO in a hybrid coupled model. Part II: Prediction with piggyback data assimilation. *Clim. Dyn.*, **16**, 35-48.
- Tang, Y., 2001: ENSO simulations and predictions using the hybrid coupled models with data assimilation, Ph.D thesis, University of British Columbia, 193pp.

- , W. W. Hsieh, B. Tang, and K. Haines, 2001: A neural network atmospheric model for hybrid coupled modeling, *Climate Dynamics*, **17**, 445-455.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge, Cambridge Univ. Pr., 484pp.
- White, W.B., 1995: Design of a global observing system for gyre-scale upper ocean temperature variability. *Prog. in Oceanogr.*, **36**, 169-217.
- Yuval, 2001: Enhancement and error estimation of neural network prediction of Nino 3.4 SST anomalies, *J. Climate*, *14*, 2150-2163.
- Zebiak, S. E., and M. A. Cane, 1987: A model El Niño-Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262-2278.

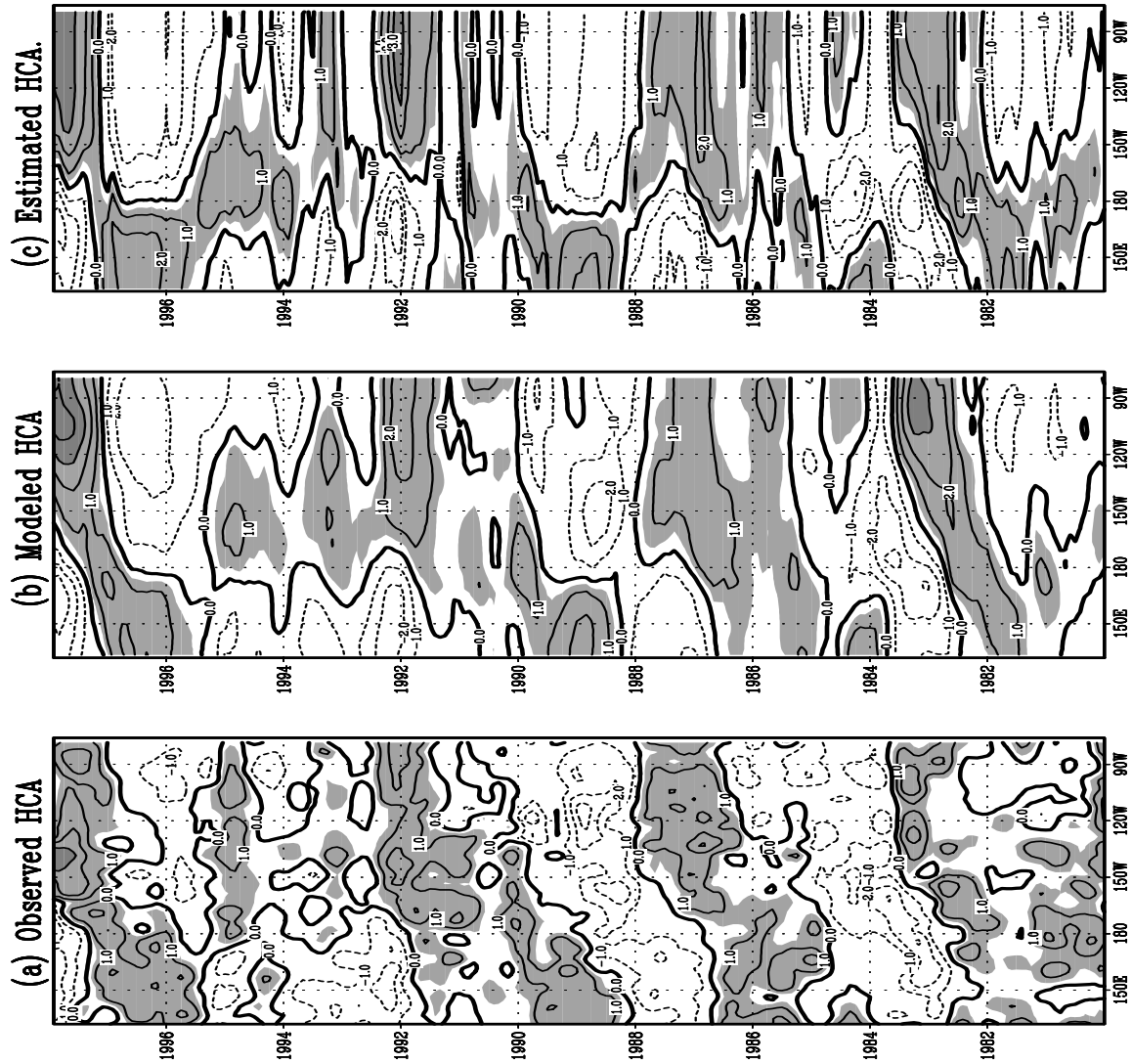


Figure 1: Time-longitude plots of the ocean heat content anomalies along the equator during 1980-1998 from (a) the observations and (b) the control experiment. (c) The HCA calculated from the prognostic variables estimated by the cross-validated ensemble NN model using (b) as predictors. Values prior to 1980 are not shown, as the quality of the observed HC is poor. Contour interval is 1°C and the values above positive 0.5°C are shaded.

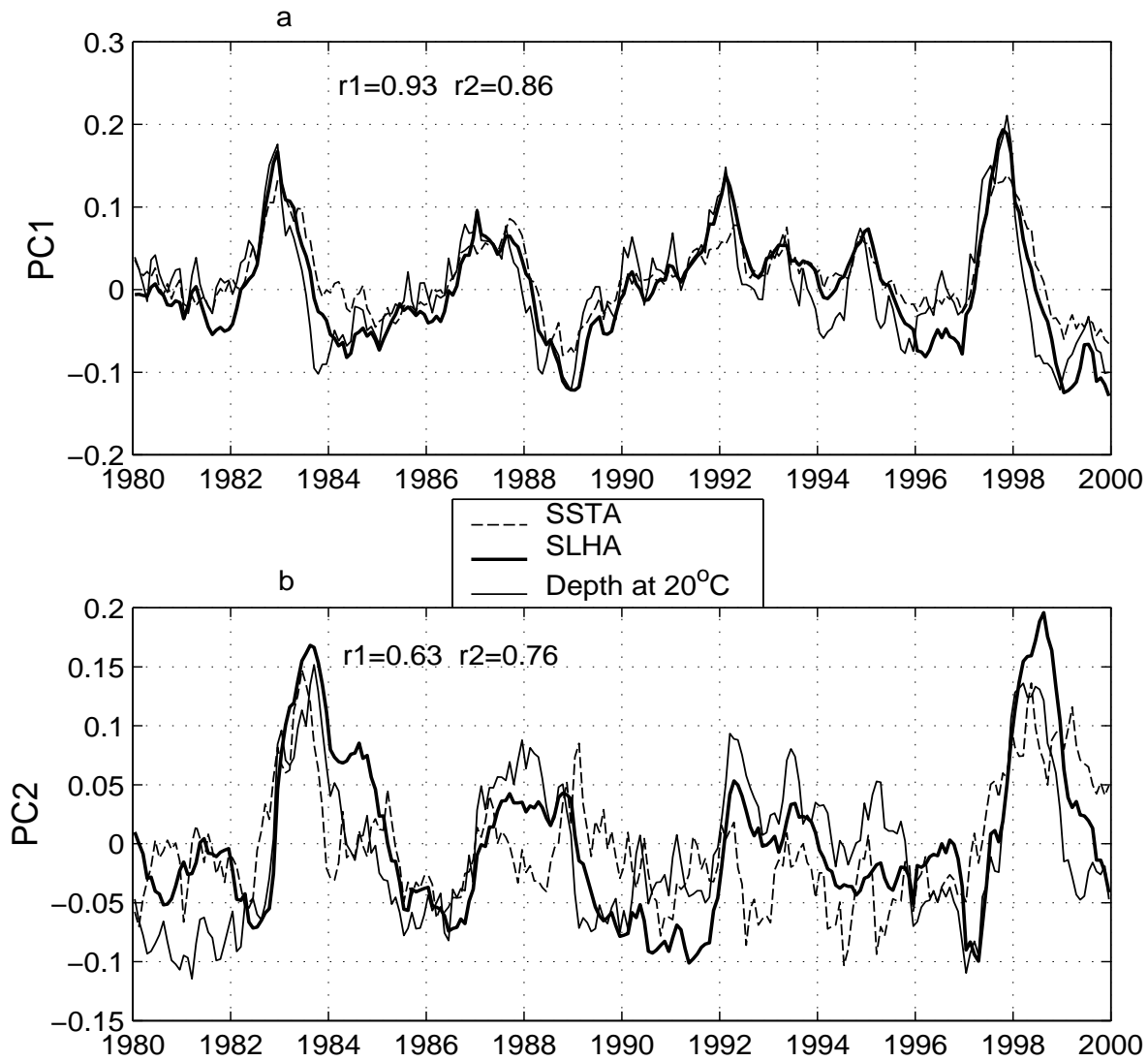


Figure 2: The time series of (a) the first principal component (PC) (i.e. EOF time coefficient) and (b) the second PC, for the NCEP SLHA (thick line), the observed SST (dashed line) and the depth of the 20°C isotherm. The first two EOF modes explain over 65% of the variance for each variable. The annual cycle has been removed prior to EOF analysis of each variable. The values of $r1$ and $r2$ are the correlation coefficients of SLHA to SSTA, and SLHA to the depth of the 20°C isotherm respectively.

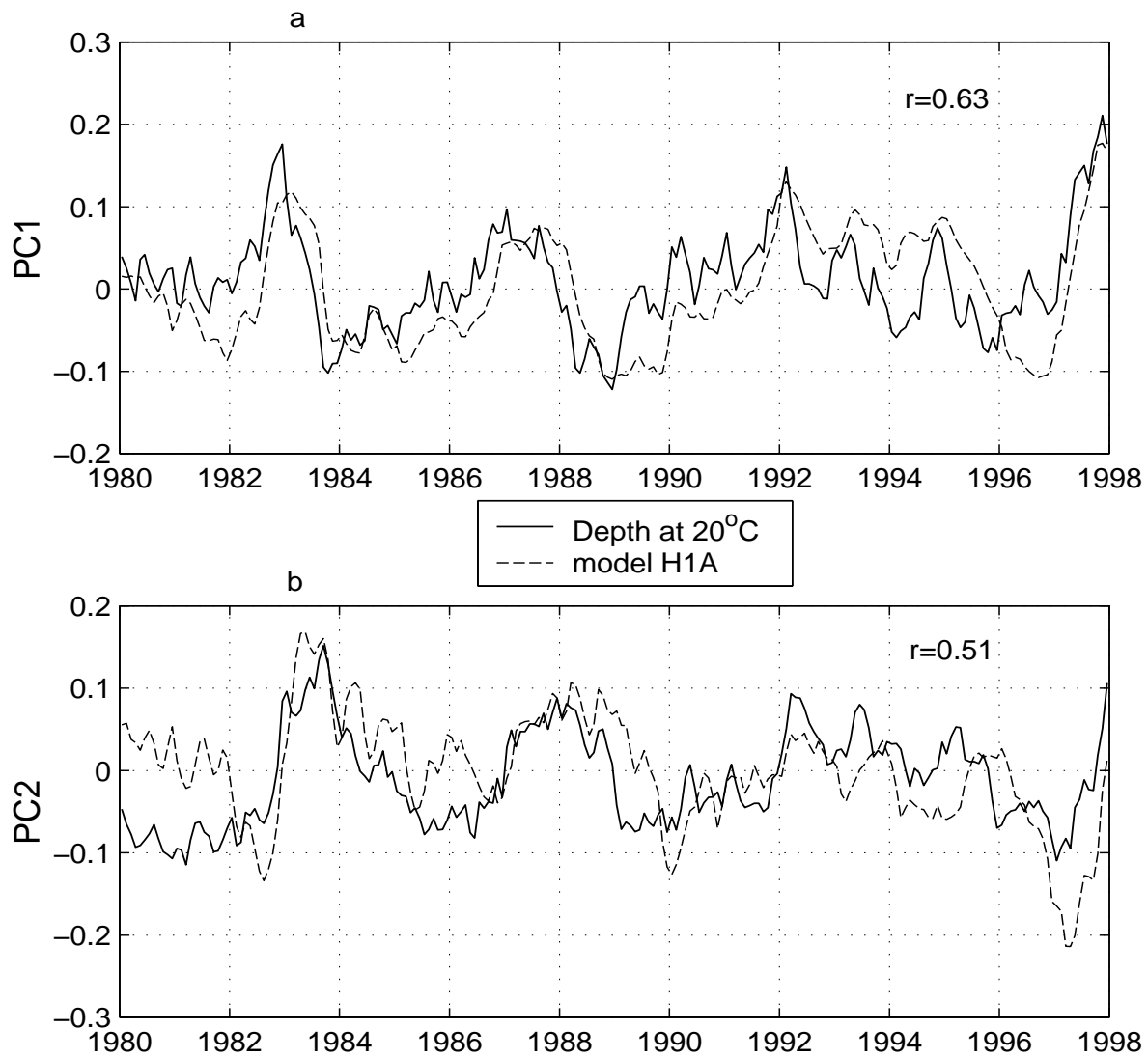


Figure 3: The time series of (a) the first PC and (b) the second PC, of the depth of the 20°C isotherm (solid line) and the modeled H1A (dash line) from the control experiment (with the ocean model forced by the FSU wind stress). The value of r is the correlation coefficient between the two curves. The annual cycle has been removed prior to the EOF analysis of each variable.

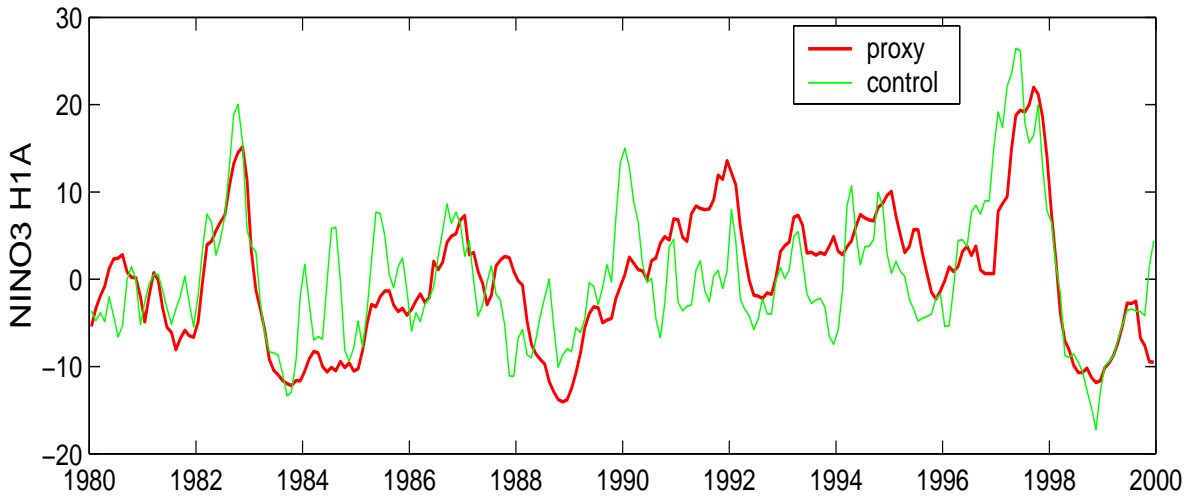
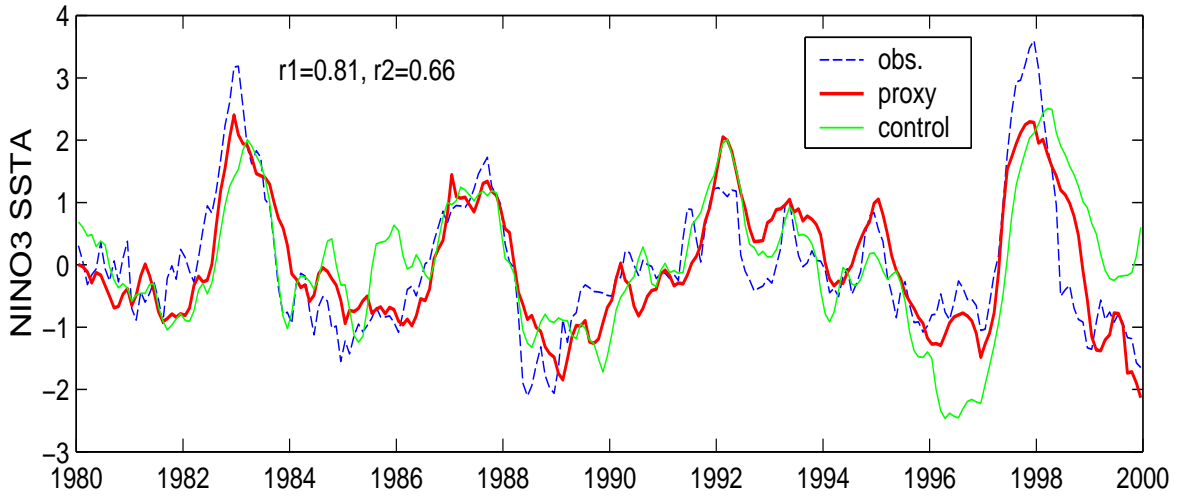


Figure 4: (a) The time series of the Niño3 SSTA index from observations (dashed line), the control experiment (thin line) and proxy SST (reconstructed from NCEP SLHA)(thick line). The correlation between the observed index and the proxy index, and between the observed and the modeled index are given by r_1 and r_2 respectively. (b) The time series of the modeled H1A (thin line) and the proxy H1A (thick line) at Niño3.

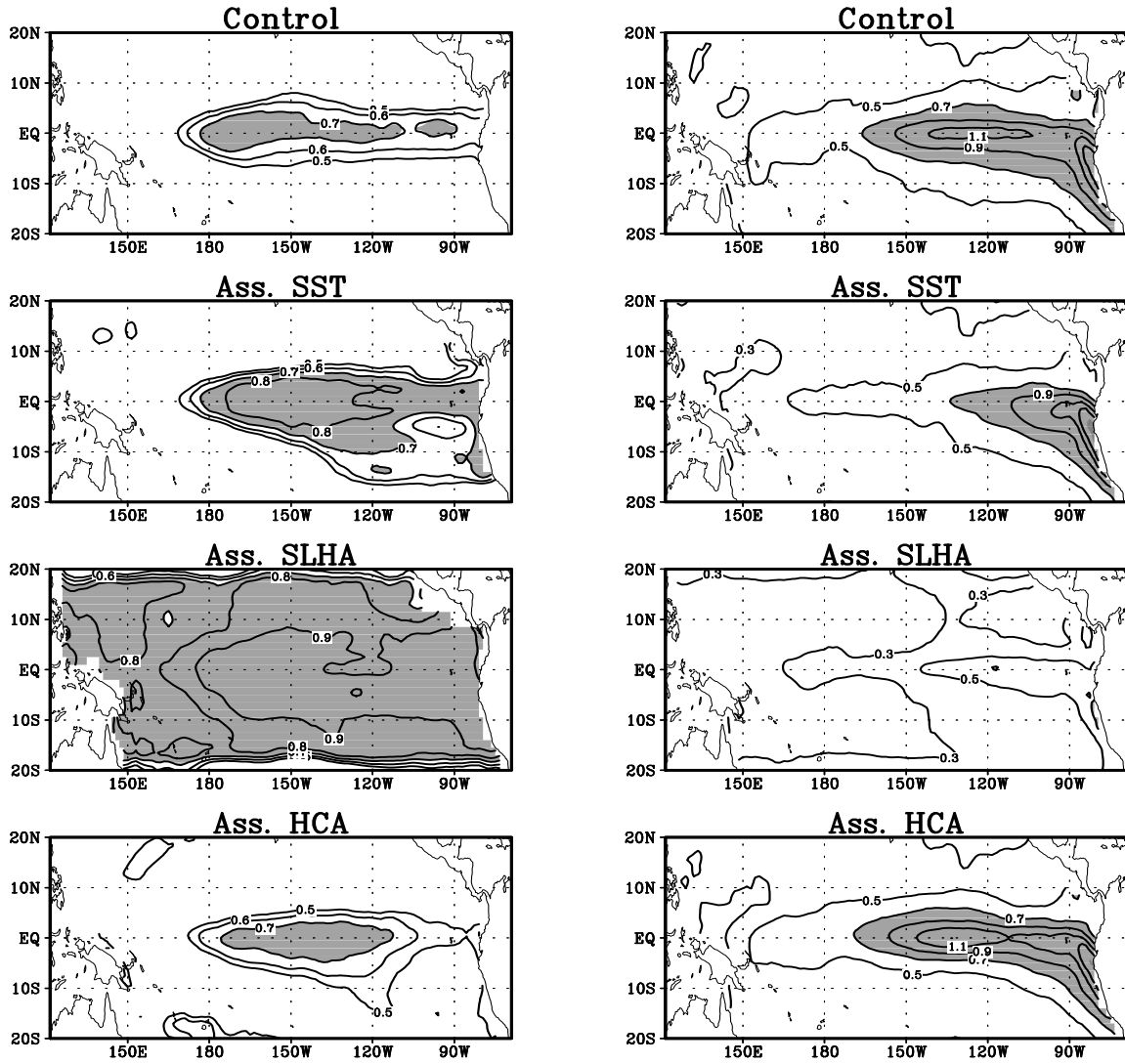


Figure 5: Correlations (left panels) and RMSE (right panels) of modeled SSTA from the control experiment and the ocean analyses with the separate assimilation of SST, SLHA and HCA relative to the observed SSTA during January 1980- April 1998. The areas with correlation over 0.7 and RMSE over 0.7°C are shaded.

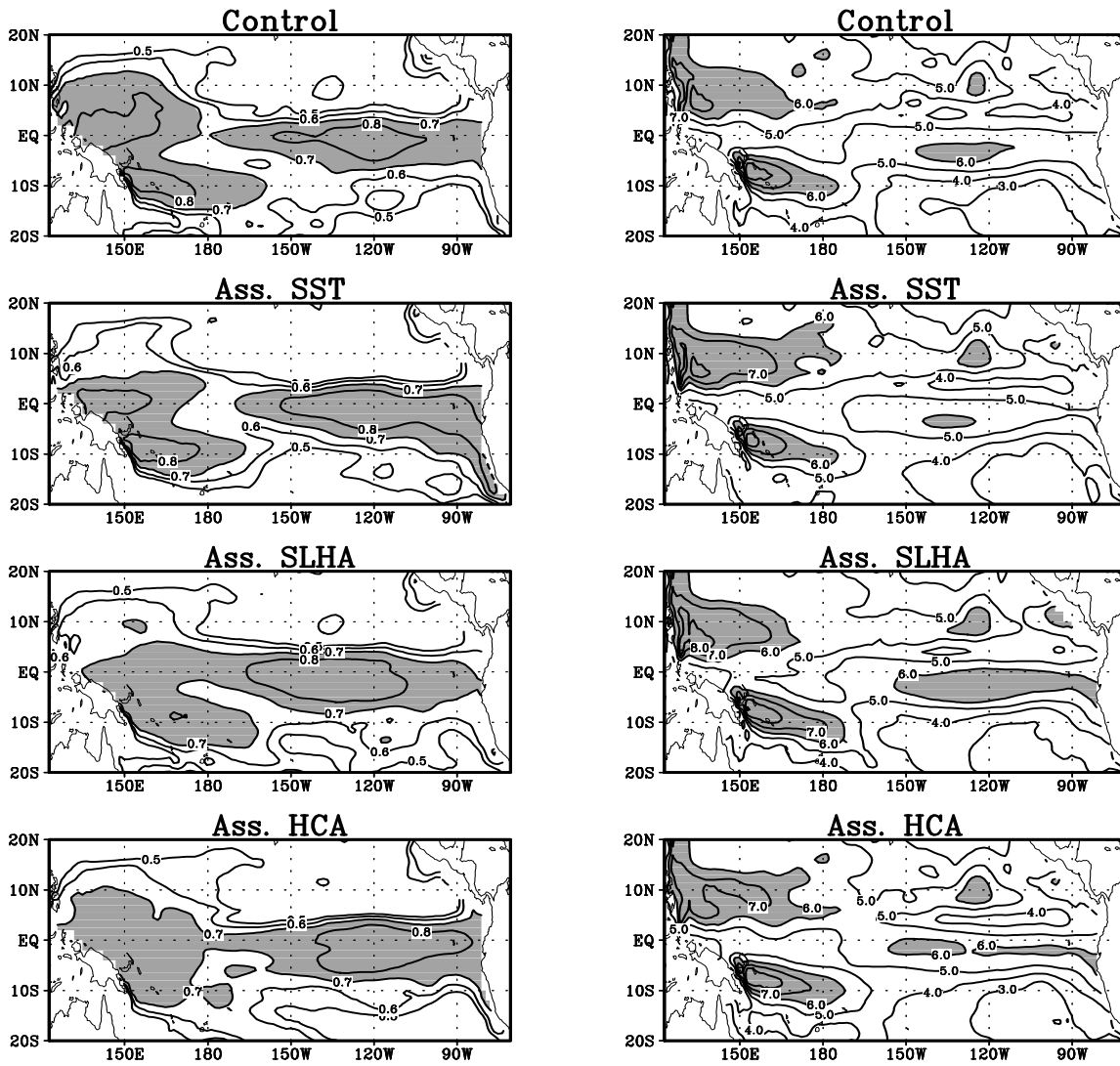


Figure 6: Same as Fig. 5 but in term of SLHA comparisons. The correlation over 0.7 and RMSE over 6 cm is shaded.

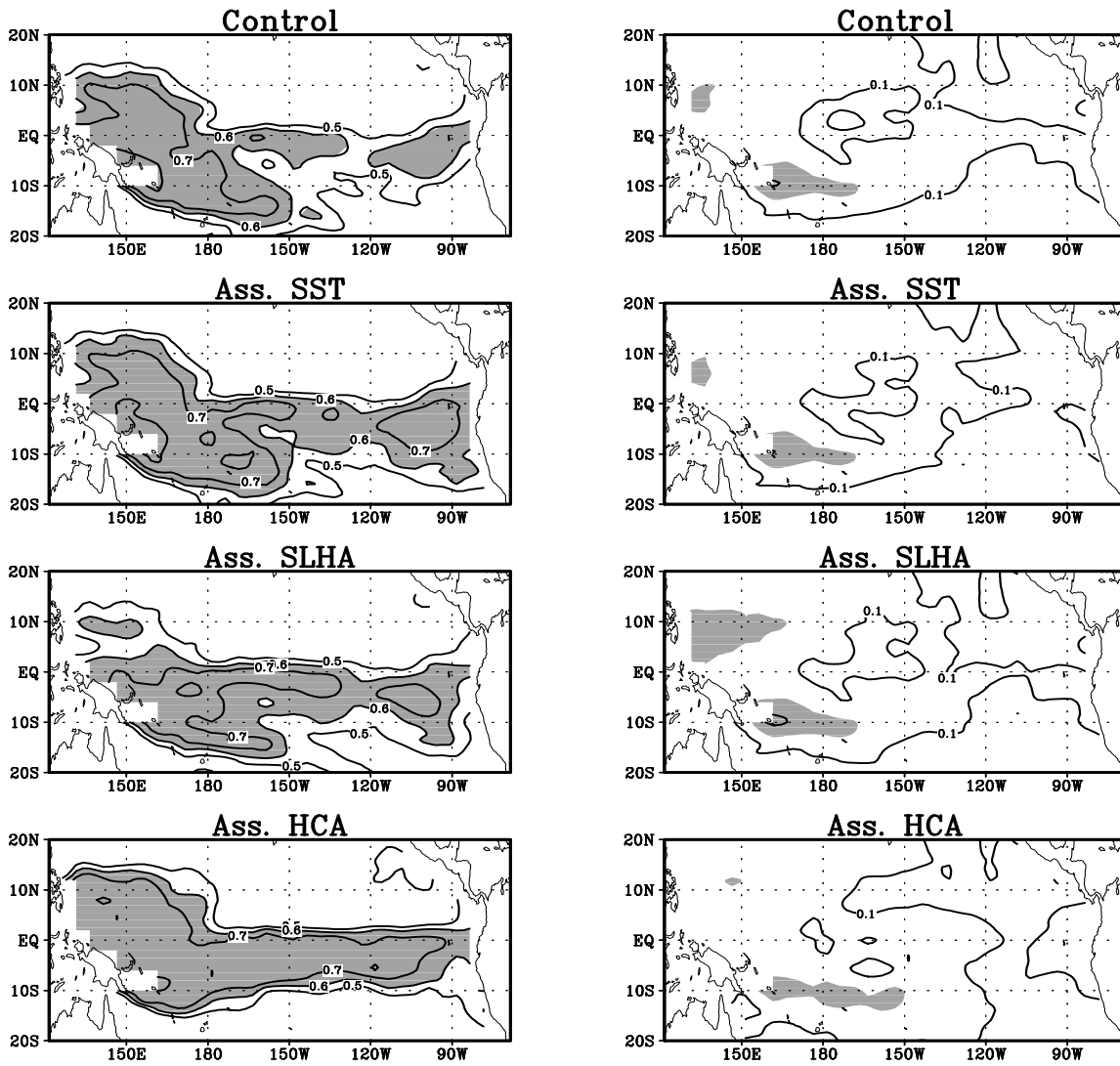


Figure 7: Same as Fig. 5 but in term of HCA comparisons. The correlation over 0.6 and RMSE over 0.15°C is shaded.

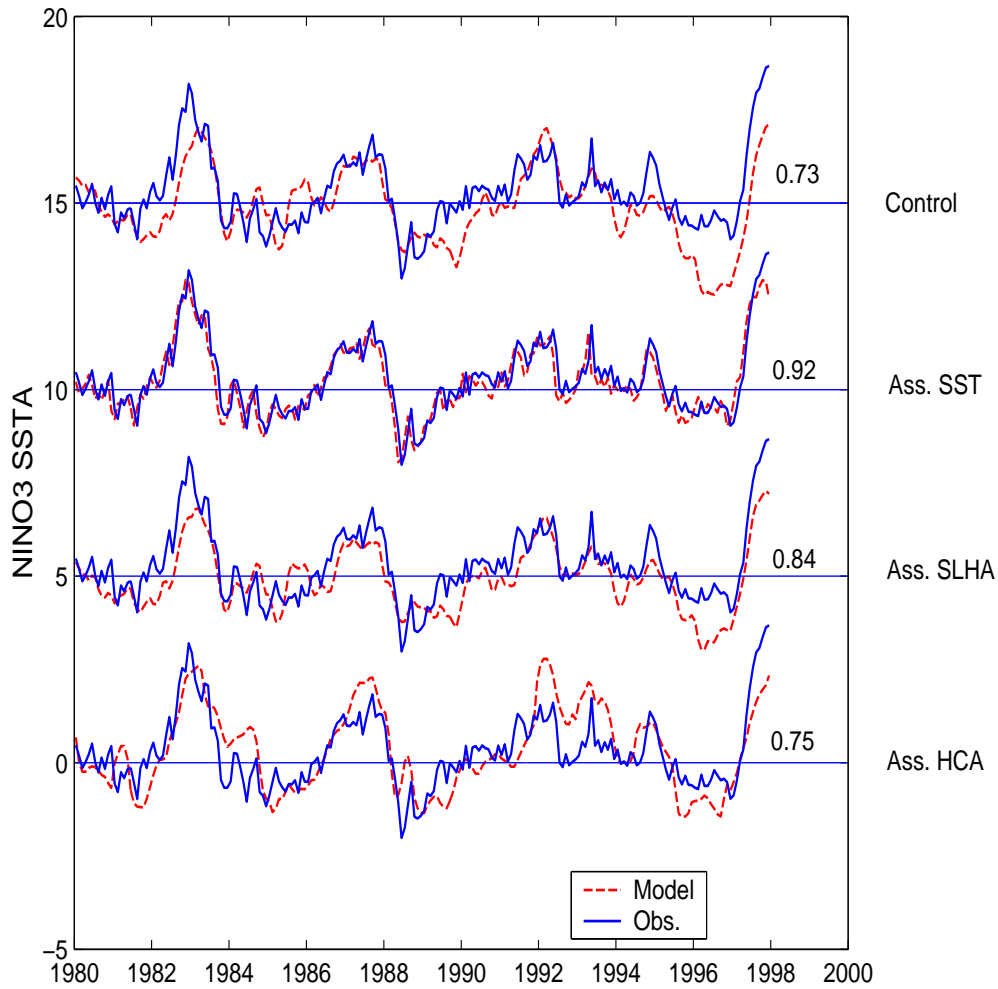


Figure 8: Time evolution of observed and modeled Niño3 SSTA during January 1980-April 1998. Solid line denotes observed SSTA and dashed line is modeled SSTA from the control experiment and the ocean analysis with the assimilation of SST, SLHA and HCA from top to bottom. The numbers on the right side are the correlations between the corresponding solid and dashed curves. For better legibility, the Niño3 SSTA for different experiments have been shifted vertically by units of 5.

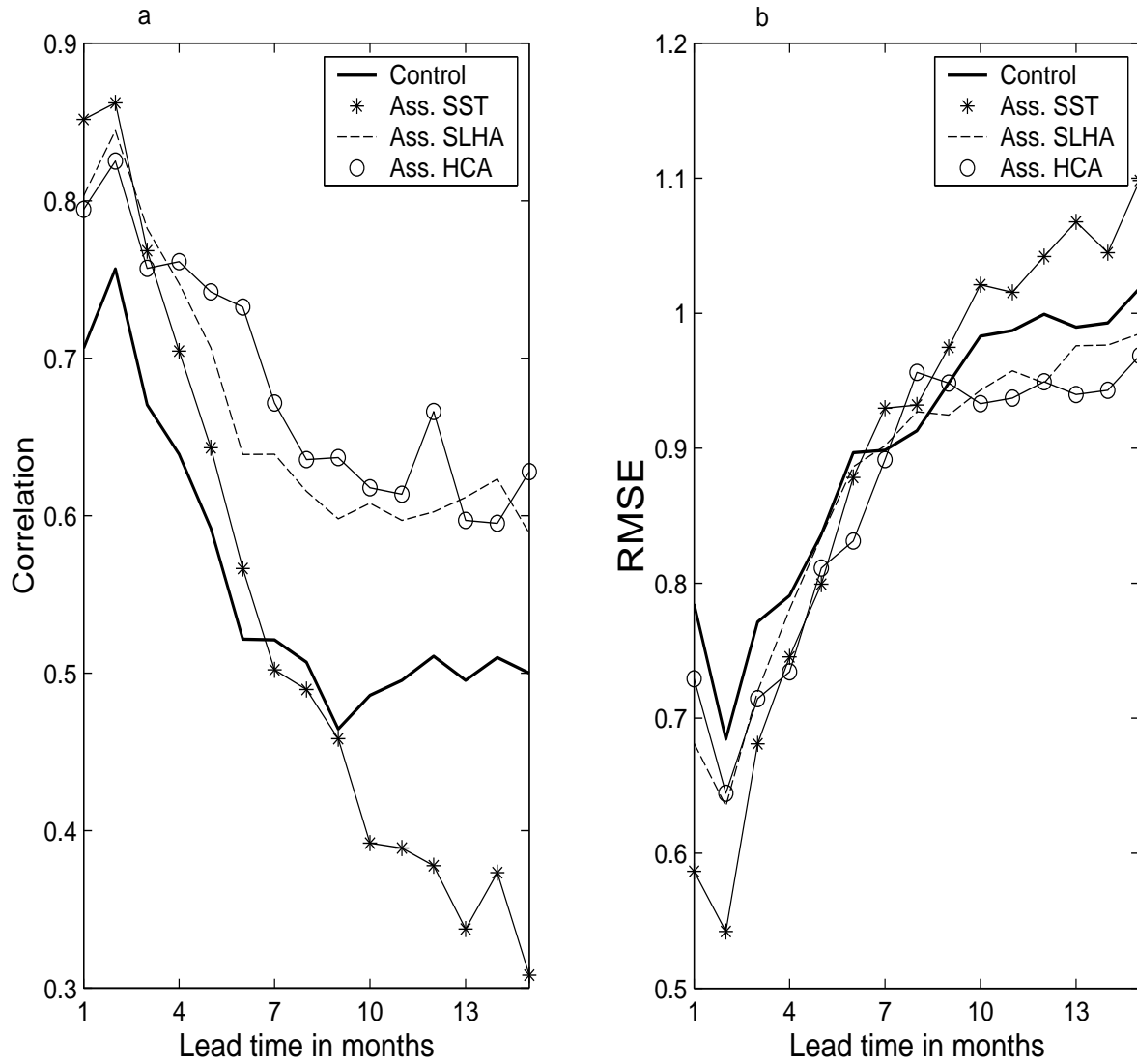


Figure 9: (a) Correlation and (b) RMSE between observed and predicted Niño3 SSTA, as a function of lead time. The predictions are initialized every 3 months from April 1980 - April 1998. The thick line is from the predictions initialized from the control experiment. The skills from the predictions initialized from the ocean analyses are shown by the asterisks for SST assimilation, dashed line for SLHA assimilation, and circles for HCA assimilation.

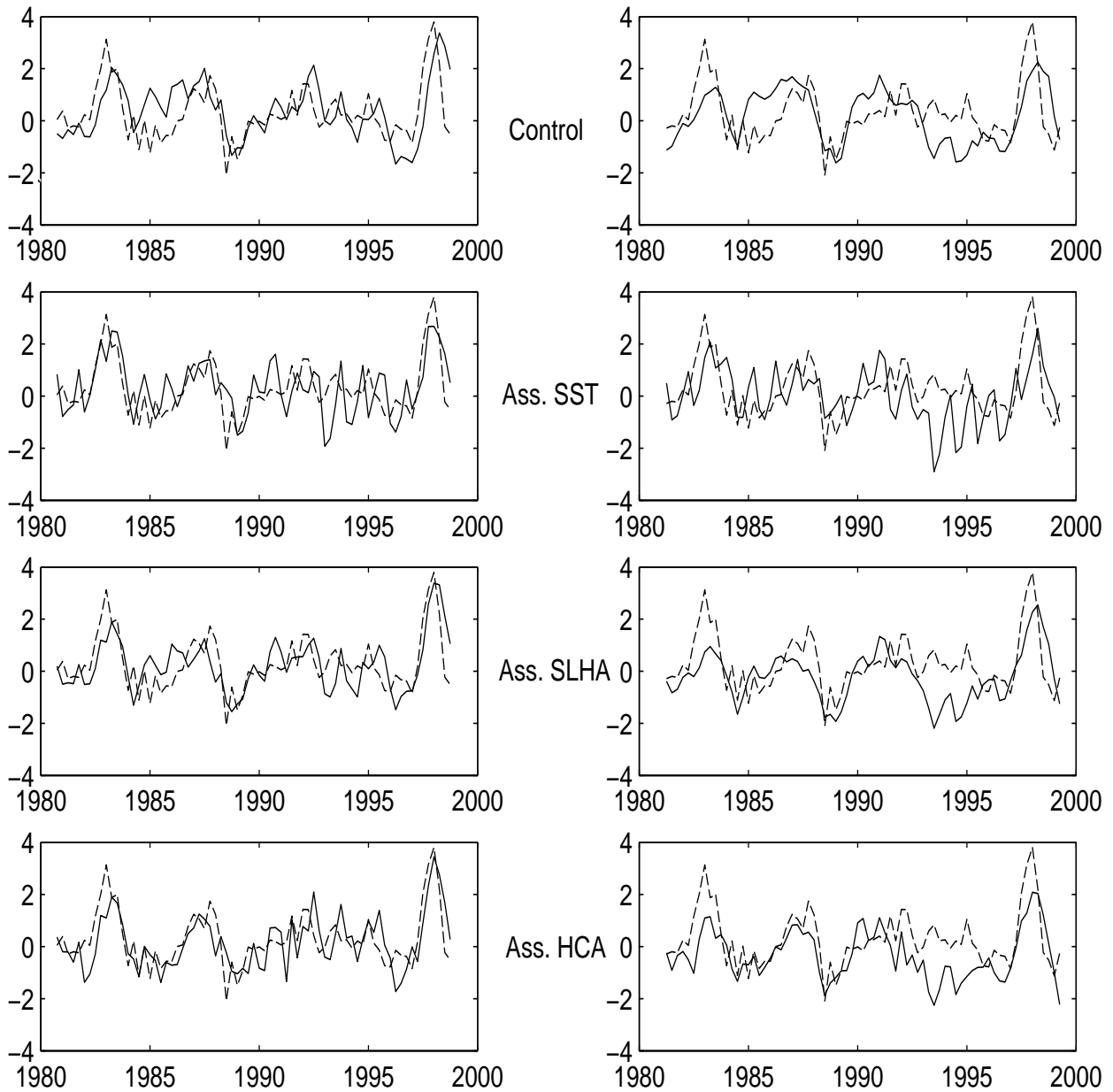


Figure 10: The predicted Niño3 SSTA (solid line) at the lead time of 6 months (left panels) and 12 months (right panels) initialized from April 1980 to April 1998, with the observed value shown by the dashed line. The predictions are initialized from the control and the ocean analyses with the assimilation of SST, SLHA and HCA (from the top to bottom).

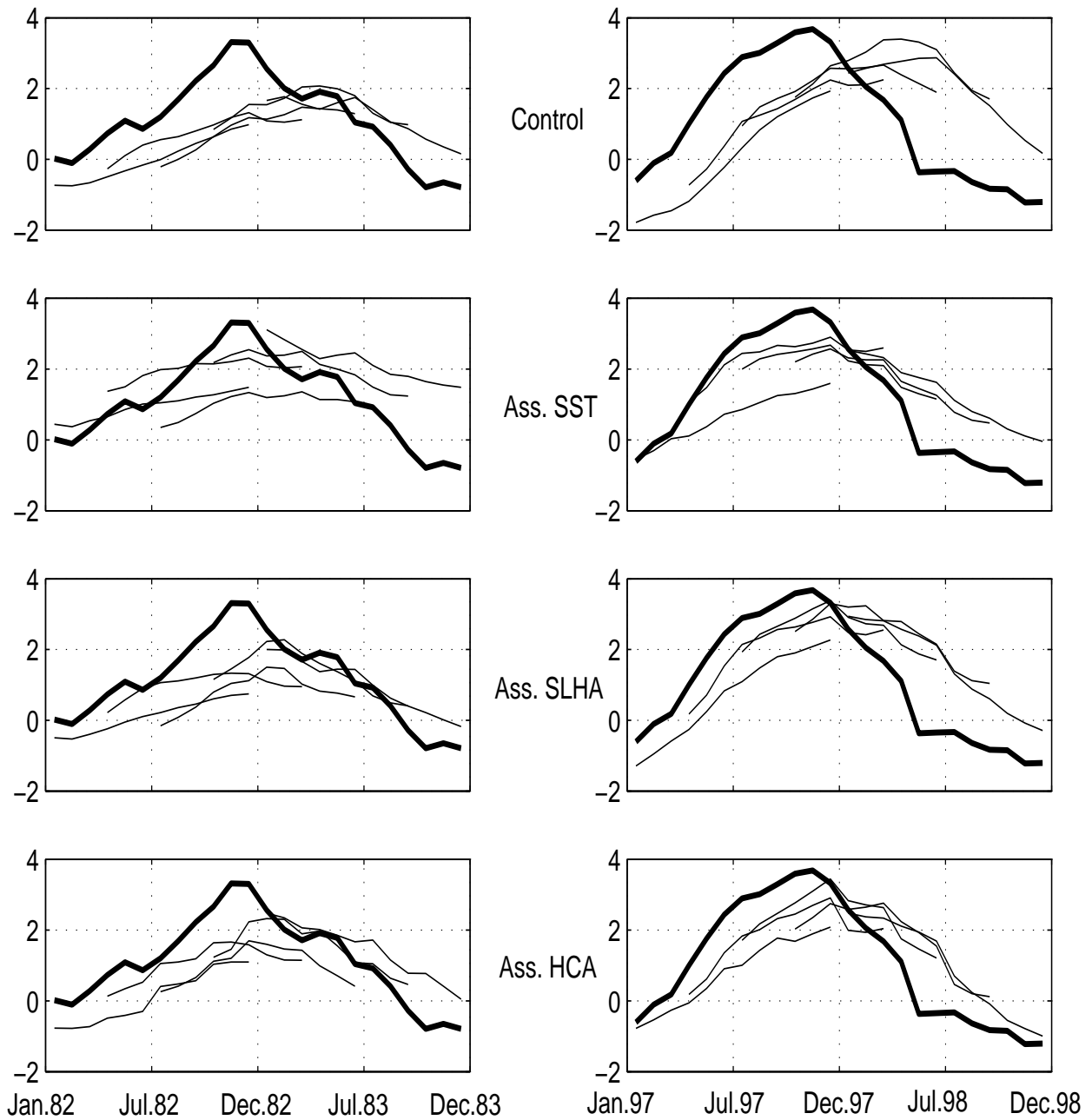


Figure 11: The predicted Niño3 SSTA (thin line) for the 1982/83 warm event (left panels) and 1997/98 warm event (right panels) initialized from the control experiment and the ocean analyses with the assimilation of SST, SLHA and HCA (from the top to bottom). The observed SSTA (thick line) is also shown for comparison. At each experiment, 5 predictions are given initialized at January 1, April 1, July 1, October 1 and the following January 1.

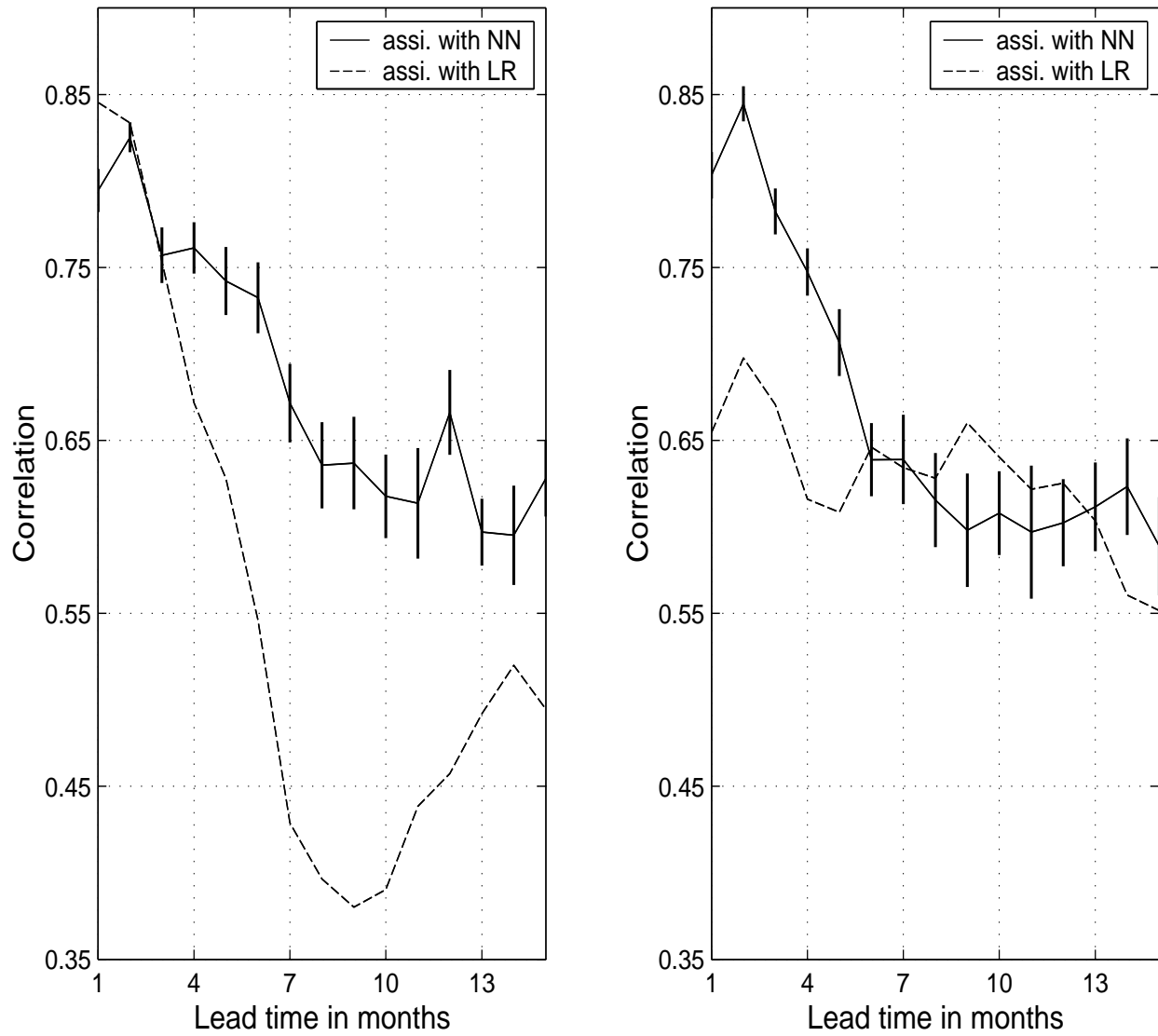


Figure 12: Correlation between observed and predicted SST anomalies in the Niño3 region, as a function of lead time, for the assimilation of HCA (left panel) and SLHA (right panel). The NN results are the solid curves; the LR results are the dashed curves. Error bars from bootstrapping are also shown for the NN results. The predictions are for the April, 1980–October, 1998 period.