A new strategy for assimilating SST data for ENSO predictions

Youmin Tang and Richard Kleeman
Courant Institute of Mathematical Sciences, New York University, New York, USA

Received 4 February 2002; accepted 10 May 2002; published 12 September 2002.

[1] With a simple 3D Var assimilation scheme, a new strategy for assimilating sea surface temperature (SST) observations has been proposed in this paper. The strategy involves assimilating two proxy data, SST and subsurface thermal data, into the ocean model. An ensemble of predictions for the Niño3 region SST anomalies (SSTA) is performed to validate the new strategy. The results show that the new strategy can effectively improve Niño3 SSTA predictions at all lead times, in particular for lead times over 6 months, and for the predictions of El Niño episodes. The prediction skills of the Niño3 SSTA attained by the new scheme can be as high as those attained by the assimilation of subsurface data and sea level height. Comparisons between two schemes of SST assimilations suggest that the impact of observations on the initializations of ENSO predictions could greatly depend on how the observations were assimilated. *INDEX TERMS:* 4504 Oceanography: Physical: Air/sea interactions (0312); 4263 Oceanography: General: Ocean prediction; 4255 Oceanography: General: Numerical modeling; 4522 Oceanography: Physical: El Niño; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation. **Citation:** Tang, Y., and R. Kleeman, A new strategy for assimilating SST data for ENSO predictions, *Geophys. Res. Lett.*, *29*(17), 1841, doi:10.1029/2002GL014860, 2002.

1. Introduction

[2] An important issue affecting the ENSO prediction skills is the initialization scheme of the prediction models. Since the middle 1990s, large efforts have been contributed to improve the initialization for ENSO prediction models through assimilating different kinds of datasets, from surface observations, subsurface in situ observations to satellite altimetry observations [e.g., *Chen et al.*, 1997; *Rosati et al.*, 1997; *Kleeman et al.*, 1995; *Ji et al.*, 2000]. It has been found that the initializations with subsurface in situ temperature observations and sea level height can significantly improve the ENSO prediction skills.

[3] The availability of data is a core issue in data assimilation. The oceanic observations are still considered spatially sparse and temporally sporadic although the ocean observing network for the tropic Pacific was vastly improved during TOGA (Tropical Ocean Global Atmosphere) [*McPhaden et al.*, 1998]. The great difficulty and expense of obtaining subsurface in situ observations also limit their use in routine predictions. As an alternative choice, satellite observations can consistently provide with us very complete surface data (such as SST and altimetry) with very high resolution. There is interest in improving ENSO predictions only through the assimilation of surface data. The comparisons between the impact of assimilating the altimetry and assimilating subsurface in situ observations on the ocean analyses and ENSO predictions have recently been addressed by some authors [e.g., *Ji et al.*, 2000]. Although SST can be obtained easily with a low cost, and has a long-term archive, few attention has been paid to initialize the ENSO predictions with it. A major reason is probably that the strategy which is usually used in the SST assimilation could not bring improvements for ENSO forecast skills as shown in *Chen et. al.* [1997], *Rosati et al.* [1997], and *Syu and Neelin* [2000].

[4] In this paper, a new strategy of assimilating SST to improve ENSO prediction skills is proposed and tested with a hybrid coupled model. The other issue addressed in this paper is how the contributions of observations to ocean states and initializations of ENSO predictions are affected by the use of observations through a comparison between two assimilation schemes. This paper is structured as follows: Section 2 briefly describes the hybrid model and the assimilation schemes in which the new strategy is introduced. Section 3 validates the new strategy by examining ENSO prediction skills, followed by the summary and discussion in Section 4.

2. The Hybrid Coupled Model and the Assimilation Scheme

2.1. The Hybrid Coupled Model and the Assimilation System

[5] The hybrid coupled model, composed of a dynamical ocean model and a nonlinear statistical atmospheric model, is identical to that in *Tang et al.* [2001]. The ocean model is one of intermediate complexity, originated from *Balmaseda et al.* [1994], but extended to six active layers in this study. The resolution of the model is $1.5^\circ \times 1.5^\circ$, covering an extension from 30°N–30°S in latitude and from 123°E–69°W in longitude. The atmospheric model is a nonlinear empirical neural network (NN) model, which predicts the contemporaneous surface wind stress field from the upper ocean heat content anomalies. The cross-validated scheme [*von Storch and Zwiers*, 1999] was used to develop the atmospheric model in order to alleviate artificial skills.

[6] The data assimilation system is the 3D Var assimilation method described by *Derber and Rosati* [1989]. Observations can be optimally inserted into the numerical models through minimizing the error between the guess field and observation. Further details about the assimilation method can be found in *Derber and Rosati* [1989].

0094-8276/02/2002GL014860$05.00

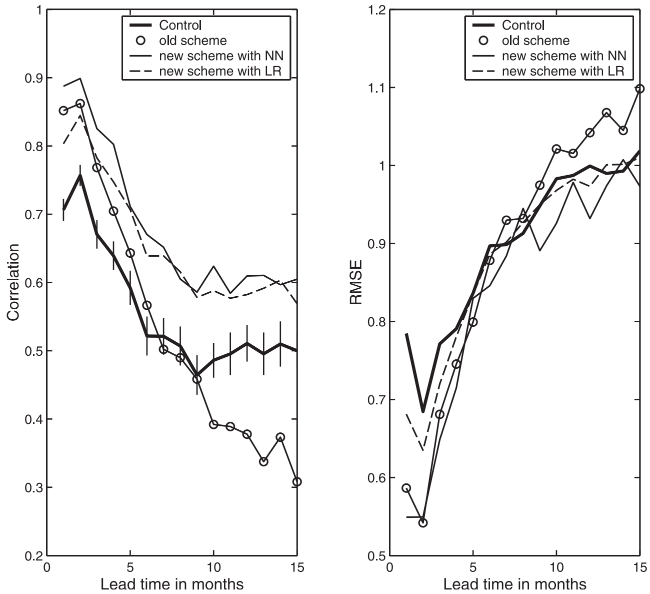

Figure 1. (a) Correlation and (b) RMSE between observed and predicted SST anomalies in the Niño3 region, as a function of lead time. The predictions are initialized every 3 months from April 1980 – April 1998. Thick line is from the prediction initialized from the control run. Circle is from the prediction initialized from the old scheme. Thin and dashed lines are from the predictions initialized from the new scheme using the nonlinear (NN) and linear (LR) relation between H1 and SST. The short line overlapped in the thick line is the error bar.

[7] Next we will discuss how to assimilate SST with the system for benefiting ENSO predictions.

2.2. A New Strategy for SST Assimilation

[8] For SST assimilation, the most common strategy is to optimally insert observations into ocean models with an assimilation scheme, i.e. directly assimilating SST observations into ocean models, because SST usually is a prognostic variable in ocean models. We used the strategy with the 3D Var system (referred to it as the old scheme hereafter) to assimilate SST observations into our ocean model during 1980–1998. With the observed error covariance of $(0.5°C)^2$, our result shows that the old scheme brings little benefit to the improvement of Niño3 SSTA predictions for lead times longer than a few months — in fact it degrades the predictions for lead times over 6 months (Figure 1). Several sensitivity experiments of the observed error covariances showed that changing the values of observed error covariances did not lead to improvements (not shown). *Chen et al.* [1997], *Rosati et al.* [1997], and *Syu and Neelin* [2000] also had similar findings when they assimilated SST into their prediction models with nudging schemes. This suggests that more considerations should be taken in SST assimilation.

[9] There are several possible reasons why the direct assimilation of SST observations by the old scheme cannot benefit ENSO predictions at long lead times. The first is that this old scheme cannot effectively correct the subsurface thermodynamical structure (thermocline). 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. 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 (not shown). These are common defects in many ocean models including OGCM [e.g., *Ji et al.*, 2000]. With the assimilation of the observed SST, the structure of the model SST is quickly forced to resemble its observational counterpart. On the other hand, however, the model adjustment is relatively slow, especially for the adjustment of thermocline, which mainly determine the variability of SST anomalies in the equatorial central and eastern Pacific. This will probably lead to some imbalance in the model physics.

[10] A new strategy is proposed to alleviate these problems. In this new strategy, the observational forcing must not be made too strong in the regions where the model SST has a significantly different structure of variance from the observations. This could be implemented in theory via a well-defined observation covariance error matrix. However, the differences of the variance structure between the model SST field and observed SST field vary irregularly in the tropical Pacific, which makes it very hard to design. Therefore we alternatively choose a more applicable approach — the assimilation of proxy SST data. This proxy SST should have the variance distribution similar to the model SST but will well keep the observational information in the temporal variability. Simply, the proxy data of observed SST is constructed through the time series of the first 3 EOF modes of the observed SST multiplying the corresponding spatial patterns of modeled SST field from the control run during 1964–1990. The proxy data keeps the observation information in time and avoid large differences of spatial distribution of variances between observed SST and model SST. Actually, if the SST assimilation is performed in EOF space (i.e., assimilating the time series of SST EOF modes), the proxy data can be approximately viewed as the SST analysis. In the new scheme, the proxy SST is used instead of observed SST in the assimilation.

[11] To more effectively adjust the subsurface structure of the model in the SST assimilation, the interface (H1) between the top layer and the second layer of the model is also simultaneously assimilated to a proxy data associated with the variations of the observed SST. The H1 anomalies of the model fairly reflect the thermocline displacement [*Tang*, 2001].

[12] The statistical relation between the modeled SST and the modeled H1 derived from the control run during 1964–1990 is sought to construct the proxy data of H1. An EOF analysis is first performed respectively for the modeled SST and H1. The statistical relations of modeled SST to H1 are extracted using only their first three EOF models, which accounted for a total of 65% of the variance for both variables. As there is not a prior reason to believe a linear relation exists between SST and H1, a nonlinear, neural network (NN) approach is used to estimate the relation [*Tang et al.*, 2001]. We also used the linear regression (LR) to relate H1 to SST, but the result is not sensitive to the linear or nonlinear hypotheses (see next section). Again, the cross-validated scheme is used to derive the statistical

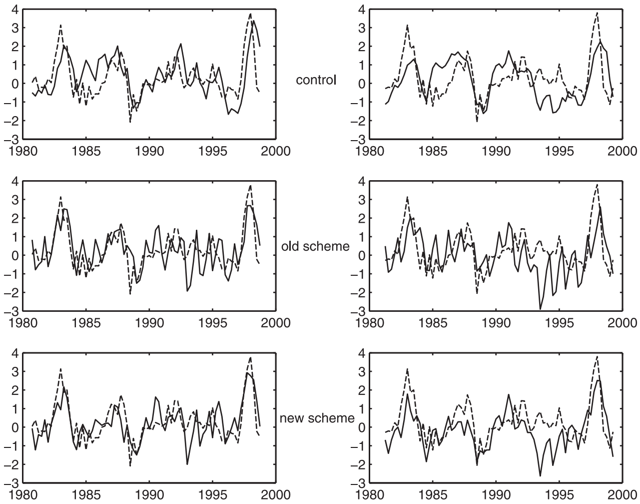

Figure 2. The predicted Niño3 SSTA (solid line) at the lead time of 6 months (left panels) and 12 months (right panels) initialized from April 1980 to April 1998. The predictions initialized from the control run and from the ocean analyses by the old scheme and the new scheme are shown from top to bottom. The dashed line is for observed Niño3 SSTA.

relations between SST and H1 as in the construction of the atmospheric model.

[13] Finally, the proxy data set of H1 associated with observed SST is obtained by using the proxy data of SST as inputs to the NNs.

[14] By this strategy, the assimilation of observed SST is performed to two proxy data sets SST and H1. The observed error covariance for SST and H1 is here set to be $(0.5^0\mathrm{C})^2$ and $(3\ \mathrm{m})^2$ respectively.

3. ENSO Prediction

[15] Next we examine the predictions by the hybrid coupled model with three initialization schemes, one from the control run, the other two from the ocean analyses with the SST assimilation by the old scheme and by the new scheme. A total of 73 prediction cases were made for each experiment from April 1980 to April 1998, starting at three months intervals (1 January, 1 April, 1 July, 1 October), and continued for 15 months. The assimilation domain is confined to the tropical Pacific, 15°S - 15°N, for all experiments. SST 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° long.

[16] Figure 1a shows the hybrid coupled model's predictive correlation skills initialized from the control run and the ocean analyses with the SST assimilations by the old scheme and by the new scheme, where the predicted Niño3 SSTA is compared against the observed values. Compared with the persistence skill (not shown), the predictions initialized from both ocean analyses beat persistence from the second month whereas the predictions initialized from the control run beat persistence from the fourth month. Generally, all suffer an initial shock in the first 1–3 months, leading to lower skill than persistence. The prediction skills initialized from the control run rapidly decline with lead time, reaching a minimum at 9 months, beyond which their skills rebound and stabilize till a lead time of 15 months— in contrast to the predictive skills initialized from the ocean analysis by the old scheme, which basically simply decline with lead time. The best predictive skills were attained with the initialization by the new scheme, with a correlation of around 0.6 for all lead times. The improvement of prediction skills initialized from the ocean analyses with the new scheme is in effect as much as that initialized by the assimilations of subsurface data and sea level height [*Tang*, 2001].

[17] As seen in Figure 1a, the prediction skills from the new scheme are not sensitive to what relation of SST and H1 is used while proxy data H1 is derived. The skills of the prediction initialized from proxy H1 with LR (dash line) is almost identical to those of the prediction initialized from proxy with NN (thin line), suggesting that there exists the approximate linear relation between the SSTA and the thermocline perturbation in the tropical Pacific, especially in the eastern equatorial Pacific where the leading EOF modes explain the big variances. The results presented below are all from the experiments using the nonlinear relation between H1 and SST.

[18] The finite sample size implies some uncertainty in the computed correlation coefficient. To determine the extent of this uncertainty we use the methodology of *Yuval* [2001]. Here a bootstrap method of sample point omission was utilized (total 73 samples), and estimates of the error are displayed in Figure 1a (short line). Clearly, the increase in the correlation skills at all lead times results from the improvement of new scheme to prediction system, rather than from the uncertainty of the finite sample size.

[19] In contrast to the improvement in the correlation, the improvement of RMSE from the new scheme is consider-

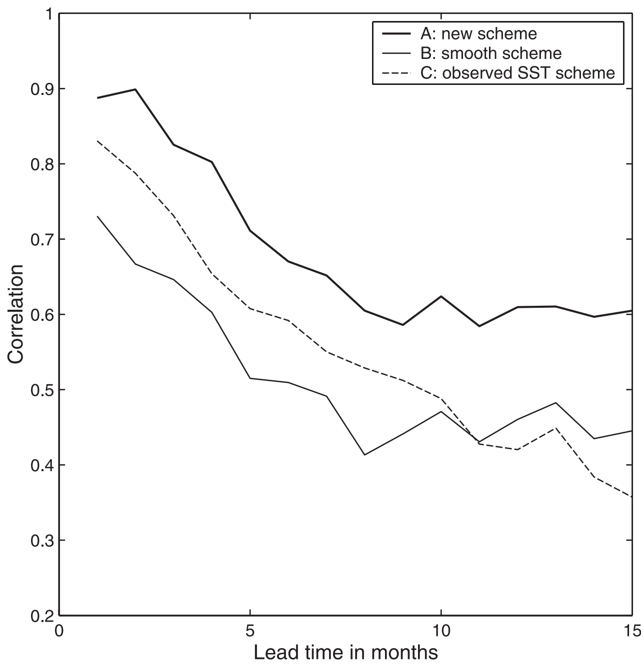

Figure 3. Correlation between observed and predicted SST anomalies in the Niño3 region, as a function of lead time. Thick line is from the prediction initialized from the new scheme. Thin and dashed lines are from the predictions initialized from two experiments (see text).

ably less impressive (Figure 1b), with the RMSE from the predictions initialized by the new scheme being slightly less than those from the control run or from the old scheme for almost all lead times. This is probably associated with the use of proxy data, which were obtained from statistical models which generally reduce the observed variance. The predictions initialized from both oceanic analyses have much smaller RMSE at lead times under 4 months than those initialized by the control run.

[20] Figure 2 shows the predicted Niño3 SSTA initialized by the three initialization schemes at the lead time of 6 months and 12 months. It was found that the initializations from the ocean analysis by the new scheme lead to the best predictions for ENSO warm events, especially in the phase prediction. The phase delay appeared in the predictions initialized by the control run were much alleviated in the predictions initialized by the new scheme, such as for the 1982/1983 and 1997/1998 ENSO warm events. However, the initializations from these ocean analyses seem to more easily produce spurious ENSO cold events, especially for the predictions initialized by the old scheme.

4. Summary and Discussion

[21] To better benefit the ENSO predictions by the SST assimilation, a new strategy was proposed and tested with a simple 3D Var assimilation system and a hybrid coupled model. Compared with the old scheme, the new scheme can better adjust the subsurface ocean, and decrease the intensity of the observational forcing in regions where there are large model biases in modeling the SST (e.g. in the western Pacific). This can be suggested by comparing the ocean analyses of the two schemes (not shown).

[22] An ensemble of predictions for Niño3 SSTA was performed to validate the new strategy during 1980–1998. The results show that the new strategy can effectively improve ENSO prediction correlation skills at all lead times, in particular for anomalous warm events. The prediction skills of Niño3 attained by the new scheme can be as high as (better than) those attained by assimilating subsurface data and sea level height for the lead times greater (shorter) than 4 months [*Tang*, 2001].

[23] A key issue in the new strategy is to construct and use the proxy data. Assimilating proxy data into ocean models for ocean analyses and ENSO predictions are often used either in situations where the observations are sparse or unavailable, or in situations where the model with a low vertical resolution cannot assimilate the available observations [e.g., *Moore and Anderson*, 1989; *Kleeman et al.*, 1995]. Both situations occurred here — there are no observations available for H1 of our ocean model, and the model vertical resolution is low. A third advantageous situation occurs, where there are large systematic differences between the spatial distribution of the variance of the observations and of the model variables. The systematic errors in the assimilation can seriously result in imbalances between the dynamical and thermal fields, i.e., the observational forcing to model SST is far faster than the adjustment of dynamical fields. The balanced state of the model at the initial time is very important to obtain a prediction skill for a longer lead time. This can be implied by an assimilation experiment similar to the new scheme, except the observed SST was used instead of proxy SST. The curve C (dashed line) in Figure 3 shows the prediction skills for the experiment, indicating the importance of using proxy SST in the new scheme.

[24] It should be noted that the new scheme also smoothes the observation and model fields by assimilating the proxy data constructed with the low order EOF modes. It is interesting to separate the smoothing contributions to the increased skill from the other main contributions (e.g the use of proxy data and subsurface adjustment). Another experiment was carried out for this purpose, in which everything was the same as in the new scheme but observed SST is replaced by model SST. The curve B (thin-solid line) in Figure 3 presents the prediction skills for the experiment, and there is little difference from the control run (Figure 1). This suggests that the smoothing contribution to the improvement of prediction skill in the new scheme is rather subtle. The major advantages of the new scheme are probably due to its ability to alleviate the imbalance between the dynamical and thermal fields when it assimilates observations into the model. This effect is absent in the old scheme.

[25] **Acknowledgments.** YT would like to thank Prof. William W Hsieh for helpful comments on a first draft of this paper. This work was supported by NSF grant 25-74200-F0960 and NASA grant 25-74200-F5044.

References

Balmaseda, M. A., D. L. T. Anderson, and M. K. Davey, ENSO prediction using a dynamical ocean model coupled to statistical atmospheres, *Tellus*, *46(A)* (4), 497–511, 1994.
Chen, D., S. E. Zebiak, and M. A. Cane, Initialization and predictability of a coupled ENSO forecast model, *Mon. Wea. Rev.*, *125*, 773–788, 1997.
Derber, J., and A. Rosati, A global oceanic data assimilation system, *J. Phys. Oceanogr.*, *19*, 1333–1347, 1989.
Ji, M., R. W. Reynolds, and D. W. Behringer, Use of TOPEX/Poseidon sea level data for ocean analyses and ENSO prediction: some early results, *J. Climate*, *13*, 216–231, 2000.
Kleeman, R., A. Moore, and N. R. Smith, 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, 1995.
McPhaden, M. J., et al., The Tropical Ocean Global Atmosphere (TOGA) observing system: A decade of progress, *J. Geophys. Res.*, *103*(7), 14,169–14,240, 1998.
Moore, A. M., and D. L. T. Anderson, The assimilation of XBT data into a layer model of the tropical Pacific ocean, *Dynamics of Atmospheres and Oceans*, *13*, 441–464, 1989.
Rosati, A., K. Miyakoda, and R. Gudgel, The impact of ocean initial conditions on ENSO forecasting with a coupled model, *Mon. Wea. Rev.*, *125*, 754–772, 1997.
Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, *J. Climate*, *9*, 1403–1420, 1996.
Syu, H.-H., and D. Neelin, ENSO in a hybrid coupled model. Part II: Prediction with piggyback data assimilation, *Clim. Dyn.*, *16*, 35–48, 2000.
Tang, Y., ENSO simulation and prediction with data assimilation, Ph.D thesis, University of British Columbia, 2001.
Tang, Y., W. W. Hsieh, B. Tang, and K. Haines, A neural network atmospheric model for hybrid coupled modeling, *Climate Dynamics*, *17*, 445–455, 2001.
von Storch, H., and F. W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge, Cambridge Univ. Pr., 484 pp, 1999.
Yuval, Enhancement and error estimation of neural network prediction of Nino 3.4 SST anomalies, *J. Climate*, *14*, 2150–2163, 2001.

Y. Tang and R. Kleeman, Courant Institute of Mathematical Sciences, New York University, New York 10012, USA.