6.5 MULTIVARIATE ASSIMILATION OF ALTIMETRY INTO AN OGCM WITH DIAGNOSTIC SEA−SURFACE HEIGHT USING THE ENSEMBLE KALMAN FILTER

Christian L. Keppenne*
Science Applications International Corporation, Beltsville, Maryland

Michele M. Rienecker
Code 971, Laboratory for Hydrospheric Processes
NASA Goddard Space Flight Center, Greenbelt, Maryland

1. INTRODUCTION

The purpose of the NASA Seasonal−to−Interannual Prediction Project (NSIPP) is to further the utilization of satellite observations for prediction of short term climate phenomena. NSIPP undertakes routine forecasts in a research framework with global coupled ocean−atmosphere−land surface models. The initial implementation has used a univariate optimal interpolation (UOI) analysis scheme to assimilate temperature data into the Poseidon quasi−isopycnal ocean general circulation model (OGCM, Schopf and Loughe 1995; Konchady *et al.* 1998).

The UOI has worked well to assimilate *in situ* temperature data from the Tropical Ocean and Atmosphere (TAO,*e.g.,* McPhaden *et al.* 1998) array into Poseidon. Nevertheless, it suffers from several limitations including the neglect of cross−field covariances and the assumption of a time−invariant, isotropic error−covariance model. In response to these limitations, a multivariate ensemble Kalman filter (MvEnKF) has been developed, implemented on a massively parallel computer architecture (Keppenne and Rienecker 2001a, KR01a herein), validated (Keppenne and Rienecker 2001b, KR01b herein) and used to assimilate TAO−temperature data into Poseidon (Keppenne and Rienecker 2001c, KRO1c herein).

In this talk, the MvEnKF is used to assimilate sea−surface height (SSH) measurements from the TOPEX/Poseidon (T/P) altimeter into Poseidon. The resulting model SSH is compared to the T/P observations. The version of Poseidon used here doesn't have a barotropic mode. So, the model SSH is a diagnostic quantity. Hence, the impact of the assimilation on the model prognostic variables is assessed by examining the depth of the $20^{\circ}C$ isotherm, a proxy for the thermocline depth across the equatorial Pacific.

2. MODEL AND METHODS

Poseidon (Schopf and Loughe 1995) is a finite−difference reduced−gravity ocean model which uses a generalized vertical coordinate designed to represent a turbulent, well−mixed surface layer and nearly isopycnal deeper layers. Here, a 20−layer Pacific basin version is used. This version has 173x164x20 grid boxes and about 2−million prognostic state variables. The details of the model configuration can be found in KR01a.

The MvEnKF is derived from a univariate ensemble Kalman filter (EnKF) algorithm initially introduced by Evensen (1994). Since then, the EnKF has been implemented for several atmospheric and ocean models of varying complexity. Keppenne (2000) and Houtekamer and Mitchell (2001) are two recent examples. The methodology has undergone many refinements along the way. For a historical perspective on the EnKF, the reader may consult Section 1 of KR01b. KR01a contains the detailed description of the MvEnKF data assimilation system and of its massively parallel implementation. The SSH experiment of Section 3 uses the same model and MvEnKF configuration used in KR01c for the assimilation of temperature (*i.e.,* 40 ensemble members distributed horizontally across 256 CRAY−T3E processors), except for the horizontal correlation scales which are shorter in this study ($l_{\lambda}=l_{\vartheta}=5^{\circ}$, where l_{λ} and l_{ϑ} are defined in KR01a).

3. APPLICATION

3.1 Experimental setup

Gridded SSH anomalies from T/P, interpolated to the model grid and averaged temporally over a five−day window, are assimilated into Poseidon using the MvEnKF. Since this initial experiment, the SSH observations have been processed along satellite tracks, but the choice to grid the data has been made here in order to more easily interpret its results. The T/P anomalies assimilated have been calculated in reference to the mean SSH over 1993−1996.

* *Corresponding author address:* Dr. Christian L. Keppenne, Mail code 971, NASA GSFC, Greenbelt, MD 20771; email: clk@janus.gsfc.nasa.gov.

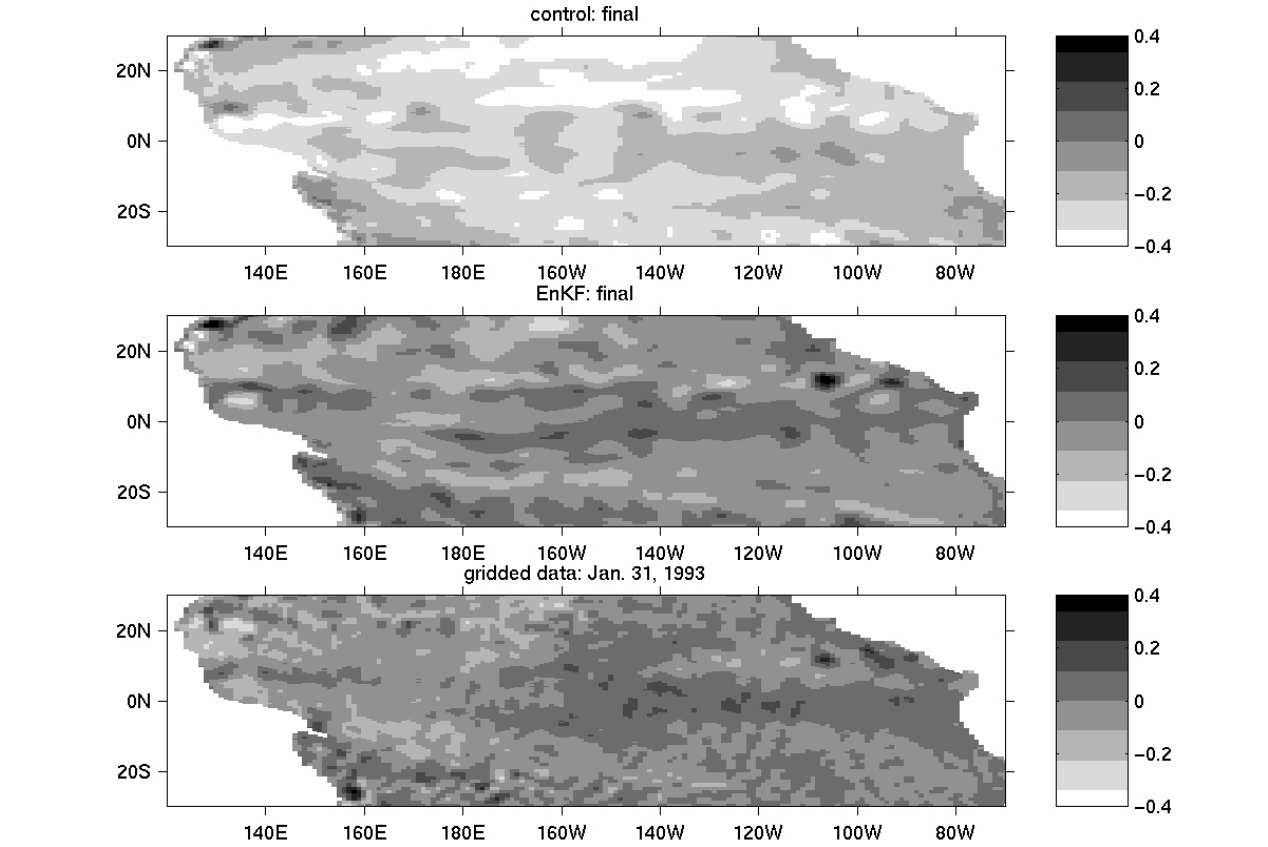

Figure 1. Hindcasts of Tropical Pacific SSH anomalies on January 31, 1993 in (top) the control ensemble without assimilation and (middle) the MvEnKF with T/P−altimetry assimilation. (bottom) Gridded T/P observations for January 31, 1993.

As the version of the model used in this experiment has a diagnostic SSH, the calculation of the temperature (*T),* salinity *(S),* zonal−current (*u)* and meridional−current (*v)* increments involves multivariate background covariances of *T, S, u* and *v* with the model dynamic height,

$$\eta(\lambda, \vartheta, t) = \frac{1}{\rho_0 \, g} \int_{\xi} b(\lambda, \vartheta, z, t) \, d\xi, \qquad (1)$$

where ρ_o is the reference density and *b,* which stands for buoyancy, is determined from the *T* and *S* distributions through the equation of state. The *h* (layer thickness) prognostic variable is not directly modified by the assimilation. Rather, it is recalculated by the model as explained in KR01b.

For reference, another ensemble run, forced with the same Spectral Sensor Microwave Imager (SSM/I: Atlas *et al.* 1996) winds and wind anomalies used to force the MvEnKF ensemble, is started from the same initial configuration as the latter, but no data assimilation is undertaken. The central forecast of this ensemble run is referred to as the control in what follows[1]. That of the MvEnKF run is simply referred to as the MvEnKF.

3.2 Sea−surface height

Since η is a diagnostic variable reflecting the model buoyancy distribution, the assimilation can only modify it indirectly through the impact on *b* of modifying the internal *T, S, u* and *v* distributions. Still, when η bears more (less) resemblance with the T/P observations in the MvEnKF than in the control, one can conclude that the assimilation has improved (worsened) the model buoyancy distribution. Therefore, it is worth looking at the impact of the SSH assimilation on η before moving on to Section 3.3.

1 A control ensemble is used rather than simply a control run to gain insight into how the background−error covariances in the MvEnKF compare to their counterpart with no data assimilation, but that discussion has been left for a future paper.

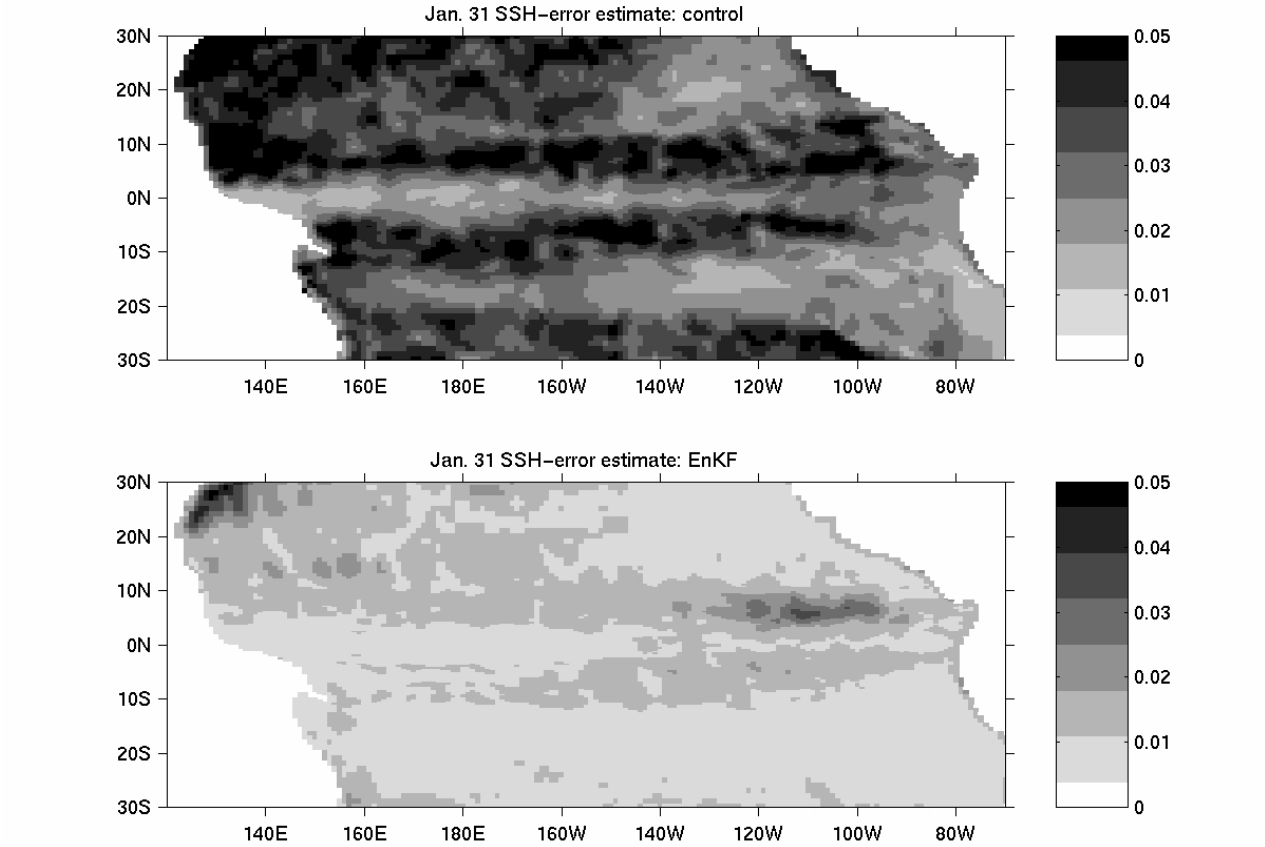

Figure 2. Estimated SSH−forecast−error for the Tropical Pacific on January 31, 1993 in (top) the control and (bottom) the MvEnKF with T/P−altimetry assimilation.

In view of the focus on seasonal prediction and on the Tropical Pacific, the plots discussed in this Section are truncated at the Tropics, although the model domain consists of the whole Pacific Ocean.

The top panel in Figure 1 shows the η forecast in the control on January 31, 1993. The middle panel shows the corresponding η forecast in the MvEnKF and the bottom panel shows the gridded T/P anomalies. The similarity between the MvEnKF and the observations is evident. To the contrary, the situation has not changed much for the control from the initial situation (not shown), in spite of one month of forcing each ensemble member with observed wind data. In the control as in the initial condition, $\eta < 0$ almost everywhere, in sharp contrast with the T/P observations[2]. Several more months of spin up are required to bring the control to the same level of agreement with the observations as that seen after a month for the MvEnKF.

The estimated forecast error for η, σ_η as inferred from the main diagonal of the background−error−covariance matrix, $\boldsymbol{P}^f$, is shown in Figure 2 on January 31, 1993 for the control (top) and the MvEnKF (bottom). Generally speaking, the amplitude of σ_η is consistent with the observations minus forecast (OMF) discrepancies (as calculated from the fields shown in Figure 1) in the control and in the MvEnKF as well. The amplitude of σ_η in the MvEnKF is about 20% of that in the control, expressing a gain of confidence in the model as more data are processed in successive MvEnKF analyses. Also, σ_η is still high in the MvEnKF after a month in the Kuroshio area. A detailed analysis of the forecasts shown in Figure 1 reveals that the OMF discrepancies remain large in that area in both the MvEnKF and the control.

2 Each ensemble member's initial state is the final state of a climatologically forced Poseidon integration (see KR01b) and only one month has elapsed since the runs were initialized on January 1, 1993.

The fact that the OMF in the MvEnKF improves over time in reference to that in the control reflects that the assimilation has impacted the model density field so that the solution, η, of (1) more closely matches the T/P data than in the control as a result of the improved buoyancy distribution.

3.3 Thermocline depth

In the upper−left panel of Figure 3, the SSH forecast along the Equator prior to the first analysis, *i.e.* $\eta(\lambda, \vartheta=0, t=0)$, is indicated by the solid line. The dashed line corresponds to the η analysis following the assimilation of the T/P data which are identified by the squares. Since η is merely a diagnostic quantity, the good match between the analysis and observations is inconsequential. What matters is whether the corrections applied to *T, S, u* and *v* result in more accurate density and buoyancy distributions subsequently.

The lower−left panel of Figure 3 provides the answer. It shows, for the last analysis at the end of January 1993, the η forecast, the analyzed η and the T/P data. In addition, the dotted line shows the η forecast from the control. The latter still differs largely from the T/P observations, as did the initial forecast. In contrast, after a month of SSH−data assimilation, η from the MvEnKF is in good agreement with the T/P observations. Therefore, the analysis and forecast at the end of January differ little from one another, in sharp contrast with the situation at the time of the first analysis. In other words, after a month, the density field has evolved so as to reduce the misfit between η and the observations. Yet, the good fit, which refers to the satisfaction of the integral constraint (1), does not require the individual layer−interface buoyancies to be accurate.

Clearly, if the *T* and S distributions have not been improved by the T/P−data assimilation, neither will *b*. Although no salinity measurements are available to examine the case of *S*, the TAO temperature data can be used to validate the *T* analyses. When the SSH data are processed, isotherms will move up or down as the layers expand or contract in response to alterations of the density field. Because the depth of the $20^{\circ}C$ isotherm, *T20,* is a simple indicator of the vertical position of the thermocline, it is used to validate the subsurface corrections made when assimilating the surface SSH data.

The two right panels of Figure 3 show the equatorial *T20* distributions corresponding to the η distributions shown in the left panels. The squares correspond to the isotherm depth inferred by bilinear interpolation of the TAO data from moorings between $2^{\circ}N$ and $2^{\circ}S$. Initially, the *T20* forecast is too deep to the east and moderately too shallow to the west. The correction resulting from the altimeter−data assimilation has the correct sign in the west. There, the analysis differs less from the (unassimilated) TAO data than the forecast does. To the east, the correction has the wrong sign as it deepens the thermocline slightly.

After a month, *T20* from the control (dots) is too shallow everywhere along the Equator. In the MvEnKF, it is on track west of $180^{\circ}W$ but still too deep to the east thereof. Also, the *T20* analysis differs little from the forecast, even to the east where the forecast−error amplitude (more precisely the root−mean−square difference between *T20* in the forecast and *T20* in the TAO data) is greater than that of the *T* increment. This behavior is expected because η has become highly accurate. Sensing this, the MvEnKF feels that the *T, S, u* and *v* forecasts require little or no correction.

Apparently, *T20* variations in the east contribute more to SSH variations in the model than in nature. The SSH errors from the control are smaller in the east than in the west. Yet, the *T20* errors are larger in the east. The corrections overcompensate leading to an excessively deep *T20* in the analysis. It appears that *T20* is embedded in the lower thermocline in the observation, whereas it is in the central thermocline in the model. This subtle relationship is the subject of an ongoing investigation.

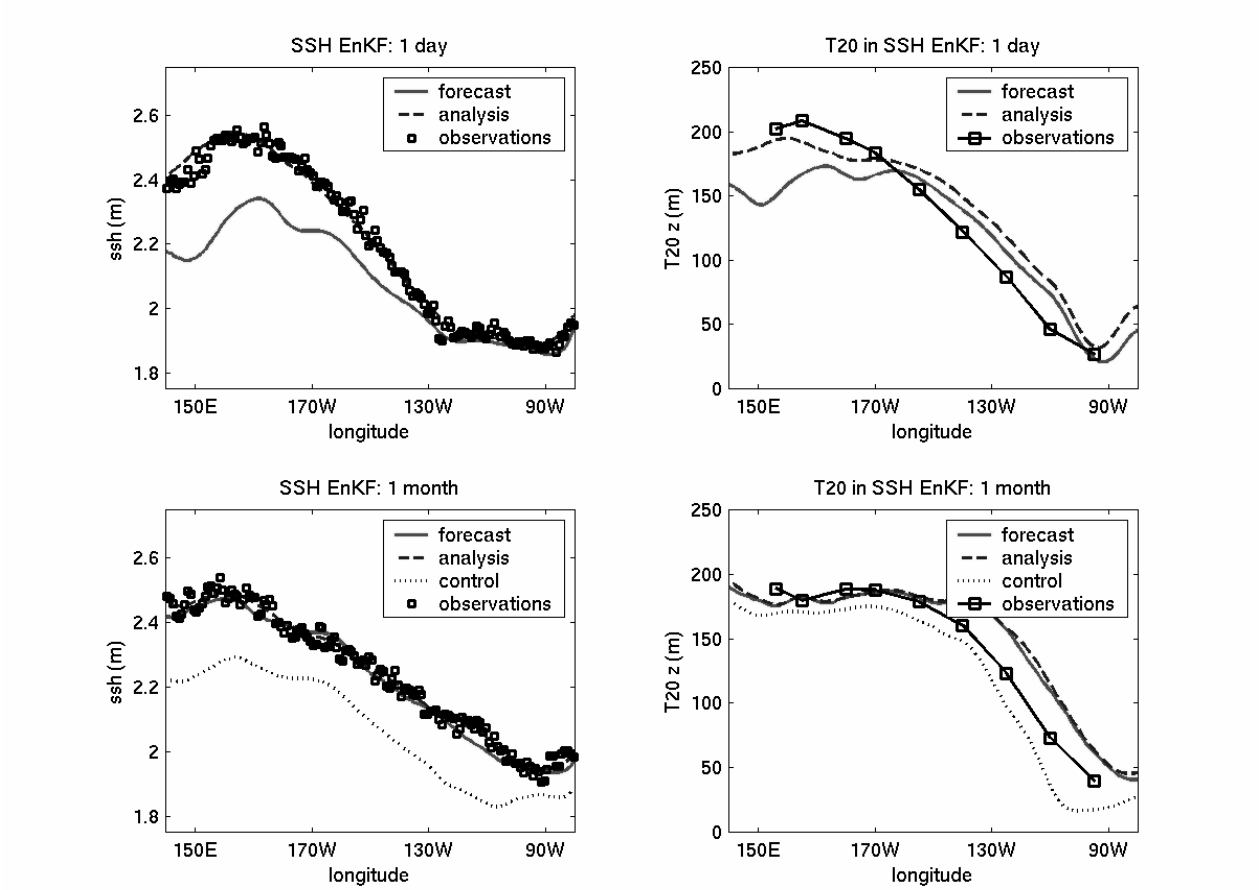

Figure 3. Left: simulated and observed SSH along the Equator (upper left) initially and (lower left) after 30 days of T/P−data assimilation. Right: simulated and observed *T20* (upper right) initially and (lower right) after 30 days. The solid and dashed lines in each panel correspond to the MvEnKF forecast and analysis, respectively. The dotted lines correspond to the control without assimilation and the squares to the observations (assimilated T/P altimetry in the two left panels and unassimilated TAO temperature data in the right two panels).

4. CONCLUSION

The objective of this study is to ascertain whether the MvEnKF can realistically be used in place of standard data−assimilation methodologies for the purpose of ocean initialization and as part of a coupled SI forecasting system. With these questions in mind, the MvEnKF is used to assimilate remotely sensed T/P−altimeter observations into the Poseidon OGCM and TAO temperature data are used for cross−validation.

The results of Section 3 and of the TAO−temperature assimilation experiments of KR01c hint that it will be necessary to assimilate data from more than one source to obtain ocean state estimates which are accurate, not only in terms of the assimilated variable, but also in terms of the other model fields. The MvEnKF is ideally suited for such applications because it dynamically estimates multivariate error statistics and considers their inhomogeneity and anisotropy in the calculation of the Kalman gain matrix. The SSH assimilation experiment of Section 3 relies on this property to update *T, S, u* and *v* even though the model has a diagnostic surface height. In this respect, the prospect of jointly assimilating satellite altimetry and vertical temperature profiles from expandable bathy−thermographs appears promising.

There are many outstanding issues. Among these, the main technical ones are related to computational cost, ensemble initialization and process−noise modeling, and are discussed in KR01a and KR01b. Another important issue is whether to perform the assimilation on constant−depth levels, as is done here and in KR01c, or within the framework of the isopycnal model formulation, *i.e.* without switching back and forth between the "layer" and "level" state−vector representations (see KR01a). Recent experiments with the MvEnKF and with the related MvOI system also developed at NSIPP have established that this choice can have significant repercussions on the outcome of the assimilation.

REFERENCES

Atlas, R., R. Hoffman, S. Bloom, J. Jusem, and J. Ardizzone, 1996: A multiyear global surface wind velocity dataset using SSM/I wind observations. *Bull. Amer. Met. Soc.*, **77**, 869−882.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi−geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, C**99**, 10,143−10,162.

Houtekamer, P., and H. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123−137.

Keppenne, C., 2000: Data assimilation into a primitive−equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971−1981.

Keppenne, C., and M. Rienecker, 2001a *Design and Implementation of a Parallel Multivariate Ensemble Kalman Filter for the Poseidon Ocean General Circulation Model.* NASA Technical Report Series on Global Modeling and Data Assimilation, Vol. 19, M. Suarez, ed., 31pp.

Keppenne, C., and M. Rienecker, 2001b: Development and initial testing of a parallel ensemble Kalman filter for the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, submitted.

Keppenne, C., and M. Rienecker, 2001c: Assimilation of temperature nto an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *J. Marine Sys.,* submitted.

Konchady, M., A. Sood, and P. Schopf, 1998: Implementation and performance evaluation of a parallel ocean model. *Parallel Comput.*, **24**, 181−203.

McPhaden, M., A. Busalacchi, R. Cheney, J. Donguy, K. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. Mitchum, P. Niiler, J. Picaut, R. Reynolds, N. Smith, and K. Takeuchi, 1998: The Tropical Ocean−Global Atmosphere observing system: a decade of progress. *J. Geophys. Res.,* **C103**, 14169−14240.

Schopf, P., and A. Loughe, 1995: A reduced−gravity isopycnic ocean model−hindcasts of El−Niño. *Mon. Wea. Rev.*, **123**, 2839−2863.