4.4 REDUCED-RANK KALMAN FILTERS: II. ASSIMILATION OF TOPEX-POSEIDON ALTIMETRY DATA INTO A REALISTIC OGCM OF THE TROPICAL ATLANTIC

Mark Buehner*, Paola Malanotte-Rizzoli and Tomoko Inui,
Dept. of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology
Antonio Busalacchi
Earth System Science Interdisciplinary Center, University of Maryland

1. INTRODUCTION

Due to the possible influence on ENSO variability, dynamical connections between the tropical and subtropical regions of the Atlantic Ocean have recently attracted considerable interest. These connections result in the subtropical winds having an impact on the equatorial sea surface temperature (SST), possibly leading to a modulation of ENSO variability. To better understand these connections, estimates of the transport between the subtropics and tropics must be improved. In this study we evaluate the ability of an efficient method for assimilating Topex/Poseidon altimetry data to improve estimates of the circulation and sub-surface thermal structure in the tropical Atlantic.

Topex/Poseidon altimetry data provided global coverage of the sea surface topography between October 1992 and April 2000. However, this instrument shares the limitation of all satellite-based instruments in that only surface information is provided. Data assimilation can provide an effective way of inferring the subsurface fields from surface observations. Sequential assimilation approaches, which are most suitable for nowcasting and forecasting applications, generally consist of a forecast step and an analysis step. During the analysis step a short-term forecast produced by the ocean model is corrected using information from a set of observations. Due to the lack of subsurface measurements, most assimiltion methods are forced to rely on a numerical ocean model to provide the surface-subsurface relationships required to make the appropriate corrections to the subsurface fields. The accuracy of the estimated subsurface fields therefore relies on how well the ocean model can reproduce these relationships.

Some studies have used a relatively simple blend of statistical and dynamical relationships to explicitly construct the extrapolation functions (*Ezer and Mellor*, 1994; *Oschlies and Willebrand*, 1996; *Cooper and Haines*, 1996). Others have used a more general approach based on the Kalman filter or extended Kalman filter (EKF). However, due to the computational expense of the EKF algorithm, some type of simplification is usually required for applications with realistic ocean models. The most severe limitation results from the need to compute the complete error covariances of the state estimates since the number of covariances scales as the square of the state dimension. To avoid this problem, most simplified versions of the EKF involve a reduction in the rank of the required covariance matrices. Several methods for choosing the reduced-dimension subspace in which the covariances are calculated have been proposed.

For the present study a reduced-rank, stationary Kalman filter was developed. We chose to use the leading EOFs calculated from the output of an unconstrained model run to define the fixed subspace for the error covariances. To simplify the assimilation procedure we first objectively mapped the altimetry data onto a regular grid, effectively making the observing array stationary. The stationarity of the observing array together with the assumption of quasi-linear dynamics suggests that an assimilation approach based on the asymptotically stationary error covariances may be appropriate.

The next section provides background information on the ocean model used in this study. In Section 3 the assimilation approach is described. The design and results of the assimilation experiments are described in Section 4. Finally, conclusions are given in Section 5.

2. MODEL OF THE TROPICAL ATLANTIC

The model of the tropical Atlantic used in this study is the nonlinear, reduced-gravity, primitive equation, sigma coordinate model developed by *Gent and Cane* (1989). The same model was used by *Verron et al.* (1999) to assimilate altimetry data in the tropical Pacific using the SEEK filter. The vertical structure consists of a surface mixed layer, employing an embedded hybrid mixing scheme, and 19 additional layers.

The model domain is 100°W to 20°E and 30°N to 30°S. The horizontal model grid has variable resolution along the meridians with the equatorial region

*Current Affiliation: Meteorological Service of Canada, Dorval QC, Email: Mark.Buehner@ec.gc.ca

better resolved (1/3° grid spacing) than regions near the northern and southern boundaries (1° grid spacing). The zonal resolution is constant throughout the domain (1° grid spacing). The model state vector consists of temperature, 2 components of velocity, and the layer thicknesses with a total dimension of 473,607. Salinity is not included in this version of the model.

The hybrid vertical mixing scheme for the mixed layer, described by *Chen et al.* (1994), combines a traditional bulk formula with a dynamic instability model. The surface forcing from wind stress and surface heat fluxes is calculated from NCEP wind speed (1958-2000) and cloud cover (ISCCP) data. Within 5° latitude of the Northern and Southern boundaries, the model fields are relaxed towards the seasonally varying Levitus climatology.

The ocean model was first integrated from a state of rest for a period of 10 years using the average annual cycle of wind stress and surface heat flux forcing to produce a fully spun-up state. The model was then further integrated over the period January 1980 to January 2000 using the actual forcing data. Both assimilation experiments begin on January 1, 1993. Although the SST varies with season, the main thermocline is generally positioned above the 18°C isotherm. The base of the thermocline corresponds approximately with the 14°C isotherm. A major underlying hypothesis in this study is the presence of a strong dynamic connection between SSH and the subsurface temperature (density) field. We confirmed the presence of this connection by calculating the correlation between SSH and the thermocline depth from the ocean model. Here we choose the 14°C isotherm as a proxy for the base of the thermocline and found a correlation greater than 0.8 in magnitude over most of the domain.

3. ASSIMILATION OF SEA SURFACE HEIGHT

The Topex/Poseidon altimetry data used in this study have undergone significant pre-processing. For the assimilation method we wish to employ it is necessary that the observation operator be stationary. Since this is not the case for the altimetry data in its original form, the data is grouped over each 10 day period and interpolated onto a uniform 1° horizontal grid covering the model domain from 25°S to 25°N. The gridded data was then interpolated in time at 3 day intervals. Due to large uncertainties associated with current estimates of the geoid, the altimetry data can only provide information on the anomalies of SSH. Therefore, the mean SSH field provided by the unconstrained model run over the period of the assimilation experiment was removed from the predicted SSH when calculating the model-observation differences. Because of this limitation, the assimilation of SSH anomalies can not be expected to provide improved estimates of the mean SSH, temperature or velocity fields.

The approach for assimilating altimetry data is similar to that developed for the first part of this study with an idealized one layer quasi-geostrophic model of the wind-driven ocean circulation. Several simplifications are made to the standard extended Kalman filter (EKF) to overcome the computational barriers encountered in applications to realistic ocean models. These include reducing the dimension of the space in which the error covariances are calculated and making the assumption that the forecast error covariances are stationary. However, the full nonlinear model is used to produce the forecasts.

Similar to *Cane et al.* (1996), we use the leading EOFs of the model state vector, including SSH, as the basis of the reduced-dimension subspace. The EOFs are calculated from the temporal variability of the unconstrained model run. For each variable type (i.e. velocity, temperature, layer thickness and SSH) the model state was first normalized by its spatially averaged standard deviation. This allowed the multivariate EOFs to be obtained using the simple Euclidean norm. The dimension of the resolved subspace is typically between $O(10)$ and $O(10^2)$ compared with the dimension of the full model state that is typically at least $O(10^5)$ for realistic models (473,607 for the tropical Atlantic model).

The model is then linearized with respect to the mean state of the unconstrained model run over the period of the experiment. With this linear propagator for the reduced-dimension subspace, the doubling algorithm (*Anderson and Moore*, 1979) is used to obtain the stationary forecast error covariances. The assumption of stationarity for the error covariances will only be valid when the observing network and the model and observation error covariances are stationary. Also, any neutral or growing modes must have a nonzero projection onto the observations and the model nonlinearities must be relatively unimportant for error growth. *Verron et al.* (1999) found little benefit from dynamically evolving the error covariances when applying the SEEK filter to the tropical Pacific.

The ocean model has a strong seasonal cycle that represents a significant proportion of the overall variability and is governed by the external forcing. It was assumed that the model accurately reproduces the average seasonal cycle. Therefore, when using model derived statistics to formulate the Kalman gain matrix, we chose to remove this cycle by filtering out the annual and semi-annual Fourier components.

The equations of the EKF required to calculate the gain matrix and the analysis and forecast error covariances are projected into the resolved subspace, resulting in the following:

$$\mathbf{K}_r = \left(\mathbf{P}_r^{-1} + \mathbf{H}_r^T\mathbf{R}^{-1}\mathbf{H}_r\right)^{-1}\mathbf{H}_r^T\mathbf{R}^{-1} \quad (1)$$
$$\mathbf{P}_r^a = \left(\mathbf{I} - \mathbf{K}_r\mathbf{H}_r\right)\mathbf{P}_r \quad (2)$$
$$\mathbf{P}_r(t+1) = \mathbf{M}_r(t)\mathbf{P}_r^a(t)\mathbf{M}_r(t)^T + \mathbf{Q}_r, \quad (3)$$

where $\mathbf{M}_r$ is the linearized dynamics in the resolved subspace, $\mathbf{H}_r = \mathbf{HE}_r$ is the linearized operator for mapping perturbations in the resolved subspace into observation space. In the expressions for the Kalman gain matrix $(\mathbf{K}_r)$ and the analysis error covariances $(\mathbf{P}_r^a)$ the cross-covariances of the forecast error between the resolved and unresolved subspaces are neglected. The matrices $\mathbf{R}$ and $\mathbf{Q}_r$ are the observation error covariances and the model error covariances in the resolved subspace, respectively. Multiplication of the innovation vector with the gain matrix $\mathbf{K}_r$ produces a correction in the resolved subspace which is used to produce the new analysis according to

$$\mathbf{x^a}(t) = \mathbf{x^f}(t) + \mathbf{E}_r\mathbf{K}_r(t)\left[\mathbf{y}(t) - \boldsymbol{\mathcal{H}}\left(\mathbf{x^f}(t)\right)\right], \quad (4)$$

where $\mathbf{y}$, $\mathbf{x^a}$ and $\mathbf{x^f}$ are the observations, analyzed state and forecasted state, respectively.

4. ASSIMILATION EXPERIMENTS

4.1 Using Simulated SSH Data

The assimilation approach applied to the tropical Atlantic model was first evaluated within a twin experiment framework. The numerical model was used to produce the "true" ocean from which the perfect observations were extracted. Then, these observations were used by the assimilation system to correct a model integration started from a different initial state. By assessing how well the assimilation system can drive the ocean state towards the "true" ocean, the effectiveness of the assimilation system can be tested independent of the quality of the numerical model.

The "true" ocean run was produced by integrating the ocean model over the three year period from January 1, 1993 to January 1, 1996 starting from a fully spun-up state. The "false" ocean run is the result of integrating the model over the same period, but starting from the incorrect initial state also used for the assimilation run. Here, the "true" ocean state from January 1, 1996 was used as this incorrect initial state. The asymptotically stationary covariances were calculated in the subspace spanned by the leading 50 EOFs calculated from the "true" ocean run sampled every 3 days. The model error covariances, $\mathbf{Q}$, were set to 0.1 times the covariances, with the average seasonal cycle removed, calculated from the "true" ocean run.

The root-mean-square (rms) error from the assimilation run and the "false" ocean run was calculated

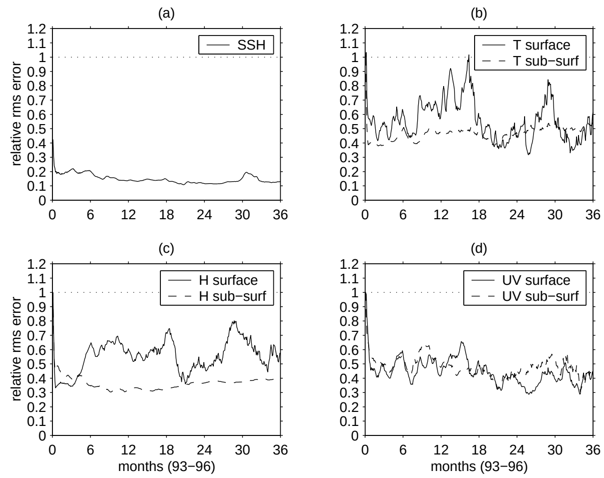

Figure 1: Relative rms error from the three year identical twin assimilation experiment. The rms error from the assimilation run is normalized relative to the run without assimilation. Error in the estimated (a) SSH, (b) temperature, (c) layer thicknesses and (d) velocity components.

separately for each variable. It was apparent that even without assimilation the rms error in the "false" ocean run steadily decreased over the three year period. Because of this it was important to measure any error reduction in the assimilation run relative to the error reduction in the "false" ocean.

Figure 1 shows the rms error from the assimilation run after normalizing by the rms error from the "false" ocean run. For all of the variables the rms error was dramatically reduced within the first 10 days, similar to the results of *Verron et al.* (1999) for the tropical Pacific. The error was reduced most for the observed variable, SSH, for which the error remains around 15% of the "false" ocean error. Below the surface the relative error for temperature is about 45% and for the SST the error ranges between 40% and 100%. The higher errors for the surface mixed layer temperature and thickness can be explained by the strong effect that surface mixing processes have on the mixed layer temperature and thickness, but which may not affect SSH significantly.

A major reason for first using the identical twin framework was to evaluate how well the information from the SSH observations can be projected onto the sub-surface thermal structure. Using the 14°C isotherm as a proxy for the base of the thermocline, the improvement in the estimate of the depth of this isotherm by assimilating SSH was evaluated. The correlation of the depth of the 14°C isotherm between the "true" ocean and "false" ocean runs and between the "true" ocean

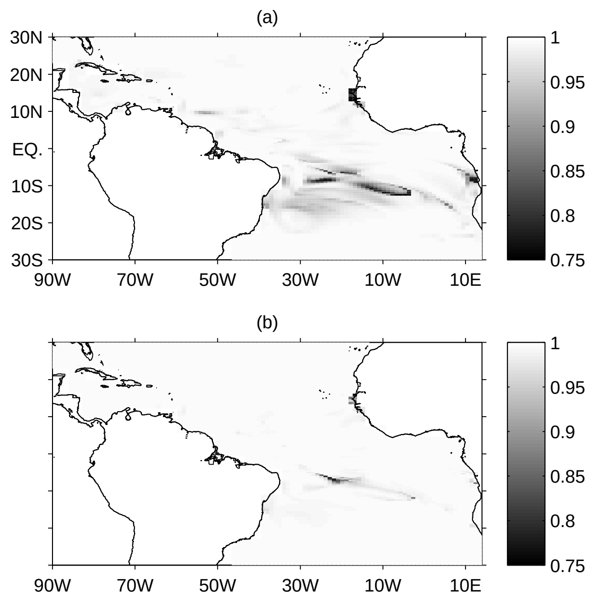

Figure 2: Correlation of the depth of the 14°C isotherm from: (a) The true vs. the no-assimilation run and (b) The true vs. the assimilation run from the identical twin experiment.

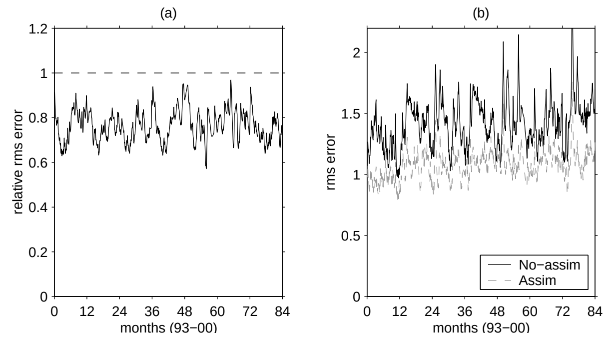

Figure 3: The rms error from the Topex/Poseidon altimetry assimilation experiment measured relative to the observations. The relative rms error after normalizing by the error from the "false" ocean run is shown in (a) and the non-dimensional rms error for the "false" ocean run and for the analyses from the assimilation run are shown in (b).

and the assimilation runs were calculated over the final year of the three year experiment. The results are shown in Figure 2a for the "false" ocean run and Figure 2b for the assimilation run. Comparing Figures 2a and 2b it is clear that the assimilation of SSH was able to effectively increase the correlation to nearly one over most of the model domain.

4.2 Using Topex/Poseidon Data

The positive results of the identical twin experiment suggest that it is possible to constrain the subsurface thermal structure by assimilating only SSH information with the reduced-rank stationary Kalman filter. However, the approach relies on many conditions which are satisfied in an identical twin framework, but may not be satisfied when using real observations, due to limitations in the ocean model to reproduce the true variability and mean state of the ocean.

The Topex/Poseidon altimetry data was assimilated over the period of January 1, 1993 to January 1, 2000. The initial state for the assimilation and "false" ocean runs was taken as the fully spun-up state on January 1, 1993 from the unconstrained model run. The asymptotically stationary covariances were calculated in the subspace spanned by the leading 75 EOFs calculated from the "true" ocean run sampled every 10 days. These EOFs accounted for 97.0% of the total temporal variability of the "false" ocean run over the period of the experiment.

The model error covariances were set to 0.01 times the covariances of the temporal variability from the "false" ocean run after filtering out the average seasonal cycle. Several values for this scaling parameter were considered and the chosen value selected on the basis that it produced the best fit of the 3 day forecast to the observed SSH.

Figure 3 shows the temporal evolution of the normalized rms error of SSH (relative to observed values) and also the nondimensionalized rms error (normalized by the model's climatological standard deviations) for the "false" ocean run and for the assimilation run. The assimilation was successful in reducing the error in SSH consistently over the full seven year period. On average, the rms forecast error of SSH was reduced by 21.5% and the analysis error by 23.6% compared to the error of the "false" ocean run.

Even though it was not possible to directly evaluate the accuracy of the subsurface fields produced by the assimilation run, we calculated the correlation between the observed SSH and the model-predicted SSH and the depth of the 14°C isotherm at two locations (20°N, 35°W and 5°N, 35°W). Both of these locations correspond to regions of high negative correlation between the SSH and the depth of the 14° C isotherm in the unconstrained model run (-0.93 and -0.65, respectively for the two locations). Therefore, the correlation of the observed SSH with the depth of the isotherm from the assimilation run (summarized in Table 1) can provide an indication of improvements made by the assimilation to the subsurface thermal structure. At the first location, the correlation with both SSH and the thermocline depth are improved significantly by the as-

Table 1:

	Location #1	Location #2
SSH "false"	−0.46	0.31
SSH Assim	0.80	0.56
Isotherm "false"	0.46	−0.22
Isotherm Assim	−0.30	−0.27

similation. The assimilation is able to change the sign of both correlations that are initially incorrect. However, at the second location that is located within the North Brazil current only a slight improvement is seen in both the estimated SSH and isotherm depth.

5. CONCLUSIONS

The goal of this study was to improve the estimates of the tropical Atlantic circulation and subsurface thermal structure produced by a reduced-gravity primitive equation ocean model by assimilating Topex/Poseidon altimetry data. A highly efficient reduced-rank, stationary Kalman filter was developed for this purpose.

By calculating the correlations from the output of the unconstrained model run a strong connection was found between SSH and the thermocline depth (14°C isotherm). This provided evidence that the subsurface thermal structure variability could be constrained by assimilating SSH anomalies alone. Results of the identical twin experiment confirmed that the chosen assimilation method can effectively recover the subsurface thermal structure and velocity field using observations of SSH. However, for the identical twin experiment several assumptions were satisfied which may not have held when using real altimetry data. These included the poor resolution of the true forecast error using model derived EOFs and error in the mean state from the unconstrained model run.

Comparing the Topex/Poseidon SSH anomalies with those from the unconstrained model run showed that the ocean model already reproduces the overall variability of SSH quite well. After assimilation, the fit to the observed SSH was improved by 23.6% and the correlation between observed SSH and thermocline depth was also improved (Table 1).

References

Anderson, B. D. O., and J. B. Moore (eds.), 1979: *Optimal Filtering.* Prentice-Hall.

Cane, M., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, 1996: Mapping tropcial Pacific sea level: Data assimilation via a reduced state space Kalman filter, *J. Geophys. Res.*, **101**(C10), 22599–22617.

Chen, D., A. Busalacchi, and L. Rothstein, 1994: The roles of vertical mixing, solar radiation, and wind stress in a model simulation of the sea surface temperature seasonal cycle in the tropical Pacific Ocean, *J. Geophys. Res.*, **99**, 20345–20359.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation, *J. Geophys. Res.*, **101**, 1059–1077.

Ezer, T., and G. L. Mellor, 1994: Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model, *J. Phys. Oceanogr.*, **24**, 832–847.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model, *J. Geophys. Res.*, **100**(C4), 6777–6793.

Gent, P. R., and M. A. Cane, 1989: A reduced gravity primitive equation model of the upper equatorial ocean, *J. Comput. Phys.*, **81**, 444–480.

Oschlies, A., and J. Willebrand, 1996: Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic Ocean, *J. Geophys. Res.*, **101**, 14175–14190.

Verron, J., L. Gourdeau, D. T. Pham, R. Murtugudde, and A. J. Busalacchi, 1999: An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation, *J. Geophys. Res.*, **104**(C3), 5441–5458.