Oceanic Data Assimilation in the South China Sea

Ping-Tung Shaw
Department of Marine, Earth and Atmospheric Sciences, North Carolina State University
Box 8208, Raleigh, NC 27695 USA; pt_shaw@ncsu.edu
Chau-Ron Wu
Center for Ocean and Atmospheric Modeling, University of Southern Mississippi
Stennis Space Center, MS 39529 USA; cwu@ssc.usm.edu
Shenn-Yu Chao
Horn Point Laboratory, UMCES, P. O. Box 775, Cambridge, MD 21613 USA

Abstract

Observational data in the South China Sea are limited. Consequently, most information on the ocean circulation is provided by numerical modeling. However, numerical simulations with imperfect estimates of the forcing fields, initial and boundary conditions, and dynamical parameters cannot achieve a high level of accuracy. To enhance realism in the simulation, the observed data are often integrated continuously into numerical models in a process called data assimilation. In this paper, three popular assimilation schemes, successive correction, optimal interpolation (OI), and a physical-space statistical analysis system (PSAS), are described. In addition, a dynamical method projecting the altimeter sea surface height to subsurface states is reviewed. Experiments using a successive correction scheme to assimilate TOPEX/Poseidon altimeter sea surface heights demonstrate some success in resolving mesoscale features in the South China Sea. The root mean square error between the simulated and observed sea surface heights is reduced by a factor of 2~3. Using a gradient boundary condition at the Luzon Strait further reduces the error, indicating the potential to improve our knowledge of transport through the Luzon Strait. The use of OI and PSAS is demonstrated in an experiment assimilating high-frequency radar-derived surface velocities for the Monterey Bay area in California. The assimilation of surface velocity data has significantly improved the model predictions of an eddy in surface circulation. These schemes could be used to improve the description of the flow field in the South China Sea.

1 .Introduction

Since the numerical modeling study of Shaw and Chao (1994), much has been learned on the basin-wide circulation and its climatological variability in the South China Sea. The seasonal reversal of a two-gyre system and the warming of the upper ocean during El Niño under a weaker monsoon wind are well established (Chao et al., 1996a; Wu et al., 1998). Figures 1a,b show the sea surface height (SSH) field derived from the TOPEX/Poseidon (T/P) satellite altimeter for summer and winter of 1993. In summer, the SSH field shows a low off the Sunda Shelf and a small high off Vietnam, forming a dipole. In addition, a high in the SSH field is present off Luzon. In winter, low SSH extends from southeast Vietnam to Luzon over the entire deep basin with local minima off Luzon and Sunda Shelf. These features are well established in numerical simulations. A basin-wide low in winter and a dipole structure in summer are, respectively, in agreement with a basin-wide cyclonic gyre and a two-gyre system with an eastward jet leaving the coast of Vietnam. However, current models with a coarse horizontal grid generally cannot resolve mesoscale features in the basin. For example, the sea level variation off Luzon, low in winter and high in summer, is not present in the numerical simulation. The north-south surface pressure gradient off Vietnam in the altimeter data is also much larger than in the numerical studies. These mesoscale features are likely related to upwelling off Luzon and Vietnam (Chao et al., 1996b). Upwelling off Vietnam has been shown in the observations compiled by Wyrtki (1961). Recent hydrographic observations (Shaw et al., 1996) lend support to the existence of upwelling off Luzon. Upwelling plays a major role in the warming of the upper ocean in the South China Sea (Chao et al., 1996a).

In view of the scarce field observations in the South China Sea, providing a detailed description of these mesoscale features and their variations is out of the question in the foreseeable future. Even models with high resolution would not help because of the poor resolution in the wind data (2.5°) and uncertainties in the transport estimates through the Luzon Strait. At present, the satellite altimeter provides the only data set covering continuously the entire basin for multiple years. A viable approach to study the mesoscale processes would be to assimilate the altimeter data into the model of Shaw and Chao (1994). In this paper, potential assimilation schemes for data assimilation in the South China Sea are reviewed.

2. Data and Model Description

The SSH from the satellite altimeter is the most promising data set for data assimilation in the South China Sea. The T/P altimeter has been operating since September, 1992. With an absolute accuracy of 4 cm, it is the most accurate altimeter system ever launched. Time series derived from the data have a sampling rate of 9 days and 22 hours and an along-track resolution of about 6 km. The

initial data processing performed at Jet Propulsion Laboratory (JPL) includes standard correction for atmospheric effects and removal of a mean sea surface and tides (Callahan, 1994). Errors from tidal aliases can further be reduced using harmonic analysis (Shaw, et al., 1999). Other data sets, such as moored data from ATLAS project (T. Y. Tang, personal communication) and XBT are possible candidates for future assimilation studies.

The model of Shaw and Chao (1994) provides an efficient tool for simulation of El Niño variations. This model solves the three-dimensional momentum, tempera-

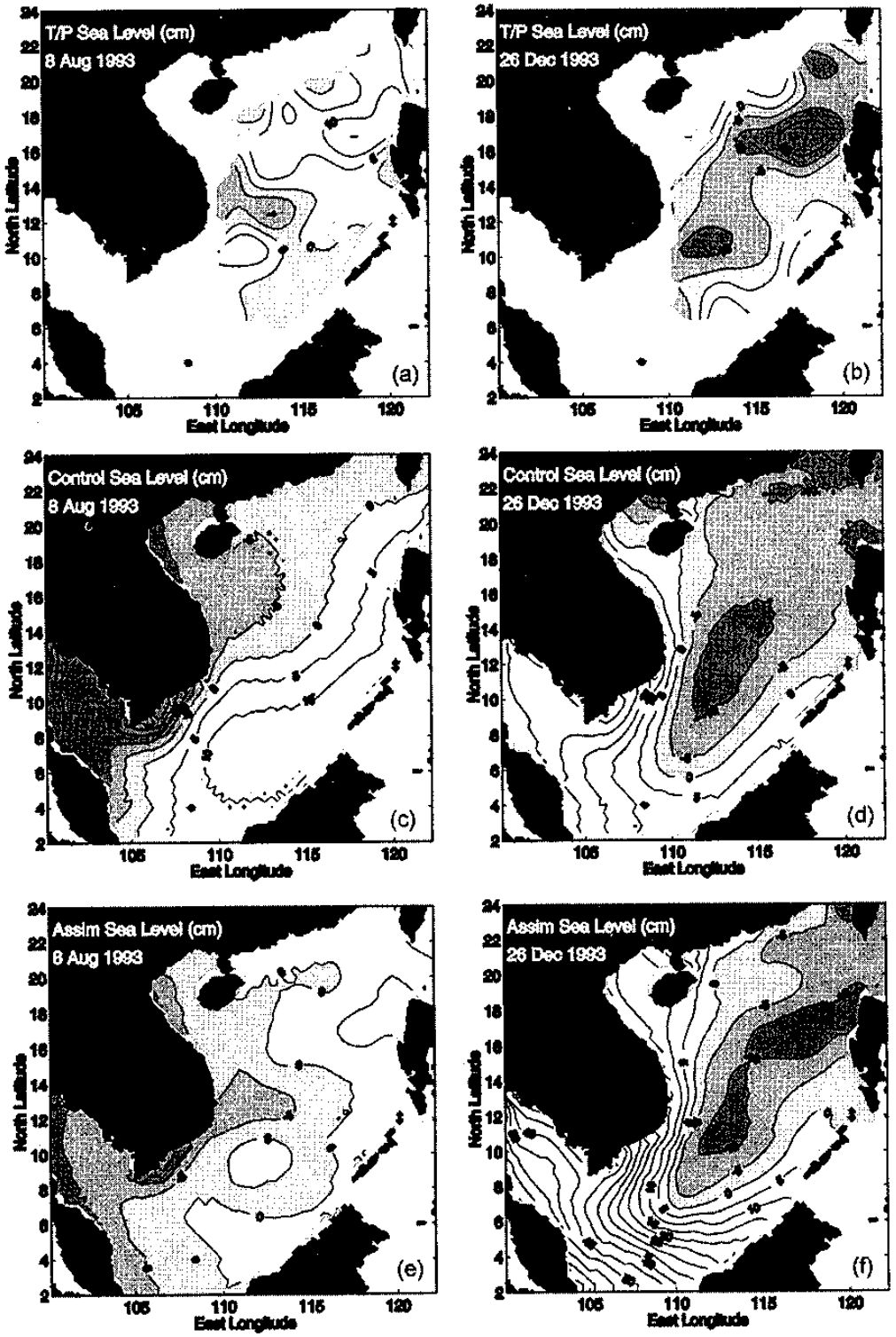

Figure 1. Sea level heights in August 1993 (left panels) and December 1993 (right panels) derived from (a, b) the TOPEX/Poseidon altimeter, (c, d) numerical simulation without data assimilation, and (e, f) numerical simulation with assimilation of altimeter data. The contour interval is 5 cm, and negative contours are shaded.

ture, and salinity equations with Boussinesq and hydrostatic approximations. The integration domain is from 2°N to 24°N and from 99°E to 124°E. The horizontal resolution is 0.4° with 21 vertical levels. At this resolution, the model could be run with time steps of 2160 s for the internal mode and 21.6 s for the external mode. The model is initialized by the January temperature and salinity fields of Levitus (1982) and is spun up by climatological forcing for one year. For the study of interannual variations, the model is forced by the NCAR/NCEP daily wind and sea surface temperatures (SST) fields (Kalnay et al., 1996). Further description of the model can be found in Shaw and Chao (1994).

3. Methods of Data assimilation

Assimilation of the altimeter data into a numerical model consists of three steps. First the SSH on the model grid is calculated from satellite SSH. Second, the subsurface oceanic states are inferred from the SSH. Finally, the analyzed surface and subsurface data are inserted into the model.

a. Interpolation of Surface Data

The successive correction method of Barnes (Daley, 1991) provides a simple approach to calculate the analyzed data on grid points. The first guess at each grid point is successively corrected using a fixed exponential weighting function. More sophisticated schemes include optimal interpolation (OI), extended Kalman filtering (EKF), and a physical-space statistical analysis system (PSAS). Among them, OI and PSAS are most promising for assimilating surface observations in the South China Sea.

Here we review the OI and PSAS schemes, following the notations and terminology used by Cohen et al. (1998). Let w^t represent a column vector of the true state of the oceanic variable w on the model grid. The forecast first guess $\mathbf{w}^f$ on the grid differs from the true value by a forecast error, ε^f:

$$\mathbf{w}^f = \mathbf{w}^t + \varepsilon^f \tag{1}$$

The observation $\mathbf{w}^o$ is related to $\mathbf{w}^t$ via a linear interpolation of $\mathbf{w}^t$ to the observation points:

$$\mathbf{w}^o = \mathbf{H}\mathbf{w}^t + \varepsilon^o \tag{2}$$

where $\mathbf{H}$ is the interpolation operator, and ϵ^o is the observation error. Under the assumption of no correlation between the forecast error and the observation error, the optimal values of the analyzed field $\mathbf{w}^a$ can be derived from the minimum variance estimation:

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{K}\left(\mathbf{w}^o - \mathbf{H}\mathbf{w}^f\right) \tag{3}$$

where $\mathbf{K}$ is the Kalman gain matrix. Let $\mathbf{P}^f$ represent the forecast error covariance matrix, and $\mathbf{R}$ the observation error covariance matrix. We have

$$\mathbf{P}^f = \overline{\left(\varepsilon^f - \overline{\varepsilon^f}\right)\left(\varepsilon^f - \overline{\varepsilon^f}\right)^T}$$

$$\mathbf{R} = \overline{\left(\varepsilon^o - \overline{\varepsilon^o}\right)\left(\varepsilon^o - \overline{\varepsilon^o}\right)^T}.$$

where the over-bar indicates averaging. In the OI scheme, $\mathbf{K}$ is calculated from

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

In the PSAS algorithm, a vector $\mathbf{q}$ is obtained first by solving the linear system

$$\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{q} = \mathbf{w}^o - \mathbf{H}\mathbf{w}^f$$

and $\mathbf{w}^a$ is calculated from

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{P}^f\mathbf{H}^T\mathbf{q}$$

The error covariance matrices $\mathbf{P}^f$ and $\mathbf{R}$ play major roles in the data assimilation schemes. They determine the "blending" of model and observational data. Despite the tremendous amount of research done on estimation of $\mathbf{P}^f$ and $\mathbf{R}$ in recent years, the problem of covariance matrix estimation remains open.

b. Vertical Projection

It is well known that using sea level heights only in the assimilation often fails in multi-layer models (e.g., Haines, 1991). Therefore, it is important to use a vertical projection scheme to propagate surface information to lower layers (Hurlburt et al., 1990). This procedure is perhaps one of the most difficult problems inherent in altimeter data assimilation. Two approaches are available. The statistical inference method uses some statistical relationships to derive changes in subsurface properties from surface variations. For example, Ezer and Mellor (1994) use pre-determined correlation functions calculated from SSH and the observed subsurface temperature and salinity fields. This approach is feasible only in regions where a large amount of data is available, such as in the Gulf Stream. In the South China Sea, few subsurface observations are available for calculating the correlation coefficients.

Another method to infer changes in the subsurface density field from SSH is to calculate the weighting function based on dynamical arguments. Cooper and Haines (1996) assume adiabatic vertical displacement in the water column when the surface height is modified. Thus, all isopycnal surfaces have the same vertical displacements Δh. Under this circumstance, water properties and potential vorticity are not modified except at the surface and bottom. Continuity in the water column is maintained by removing or adding water at the surface and bottom. For $\Delta h > 0$, light water of thickness Δh is removed from the surface, and the same

amount of dense water is added to the bottom. While for $\Delta h < 0$, light water is added to the surface, and dense water is removed from the bottom. The amount of vertical displacement Δh is related to the change in SSH by assuming that pressure at the bottom is not changed by the adjustment. Let the surface pressure change associated with difference in altimeter SSH (η^o) and model SSH (η^f) be $\Delta p_s = \rho_s g(\eta^f - \eta^f)$, where ρ_s is the mean density of sea water at the surface, and g is the gravitational constant. The process of removing light water at the surface and adding dense water at the bottom produces a pressure change $\Delta p_b = g\,\Delta h\ (\rho_b - \rho_s)$, where ρ_b is the water density at the bottom. Requiring $\Delta p_b = -\Delta p_s$ we have

$$\Delta h = \frac{\rho_S}{\rho_S - \rho_b}\left(\eta^o - \eta^f\right)$$

The associated changes in temperature and salinity are

$$\Delta T = \frac{\partial T}{\partial z}\Delta h \quad \text{and} \quad \Delta S = \frac{\partial S}{\partial z}\Delta h.$$

The vertical derivatives in the above equations can be calculated using cubic splines. A convective adjustment is used to remove unstable stratifications.

c. *Forecast Step*

In this step, ΔT and ΔS are incorporated in the temperature and salinity equations using a Newtonian relaxation scheme (Anthes, 1974). In addition, the velocity field is relaxed to the geostrophic velocity associated with the density field. The latter procedure avoids dynamical imbalance in the velocity field and is applied in deep waters only. There are several reasons for this. First, the assumption of no pressure change at the bottom may not hold in shallow waters. Second, the altimeter data in shallower waters are less reliable. Finally, vertical projection in shallow waters may lead to outcropping of isopycnals.

4. The Assimilation Study of Wu et al. (1999)

Assimilation of the altimeter data into the model of Shaw and Chao (1994) has been carried out by Wu et al. (1999). In this study, the altimeter SSH is interpolated onto the model grid (0.4°) with a sampling interval of 10 days, using Barnes successive correction method. The surface height change at each grid point is then projected to the subsurface changes in temperature and salinity, using the dynamical projection method of Cooper and Haines (1996). Data assimilation is performed over a four-year period from January 1, 1993 to December 31, 1996. The model is forced by the NCAR/NCEP daily wind and sea surface temperatures (SST) fields.

The seasonal patterns obtained from simulations with and without assimilation of the altimeter data are shown in Figure 1. The highs and lows off Luzon and central Vietnam in summer in the altimeter data (Figure 1a) are not well resolved in Figure 1c but are reproduced in Figure 1e. The two local minima in winter (Figure 1b) are also shown in Figure 1f but not in figure 1d. Therefore, data assimilation could resolve mesoscale features in the South China Sea, but the range of sea level variation is smaller than that in the altimeter SSH even with data assimilation. It is likely that poor resolution in wind fields in the South China Sea reduces both the wind intensity and spatial variability (Wu et al., 1998). Overall, model simulation with data assimilation has improved the SSH field considerably compared to the control experiment without data assimilation.

Quantitatively, the improvement in model simulation is measured by the root mean square (RMS) error ϵ between the model simulated SSH (η_i^f) and the observed T/P altimeter SSH (η_i^o):

$$\varepsilon = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\eta_i^f - \eta_i^o\right)^2\right]^{1/2}$$

where N is the number of grid points in water deeper than 1000 m. Figure 2 shows a profound annual cycle in ϵ in the control experiment (marked by "W"). Peak RMS errors appear in January-February and in July-August. The error is

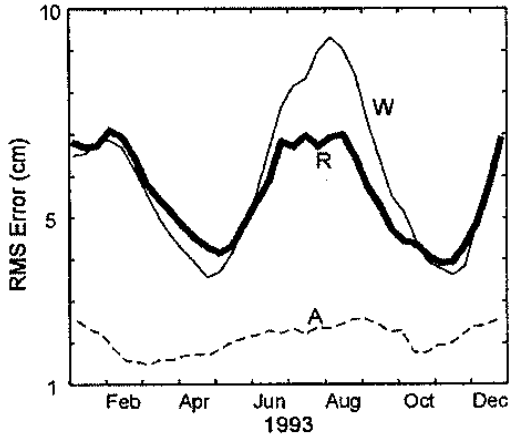

Figure 2. The RMS error in 1993 between the T/P SSH and model-produced SSH using (a) Wyrtki's transport estimates without data assimilation (thin solid line W), (b) radiation boundary conditions without data assimilation (thick solid line R), and (c) Wyrtki's estimates with data assimilation (dash line A).

smaller in transition months (March-April and October-November). It seems that large RMS errors occur in months when the wind-driven currents are strong. However, currents are stronger in winter than in summer, but the RMS error is larger in summer than in winter. Factors other than winds may also contribute to the RMS error. Figure 2 shows that the RMS error is reduce by a factor of 2 to 3 in the experiment with data assimilation (marked by "A"). Since the observed SSH field is incorporated through temperature and salinity variations, smaller RMS values also

demonstrate that the dynamical approach of projecting sea level information onto the vertical density structure is successful.

As mentioned earlier, depth-integrated transports at the open boundaries are fixed to Wyrtki's (1961) bimonthly estimates. Specification of the boundary transport is a potential source of the RMS error. To test this idea, an experiment without data assimilation has been conducted using a different open boundary condition for the transport. Instead of Wyrtki's transport estimates, vanishing normal gradients of the barotopic flow are used in the Luzon Strait, Taiwan Strait and the Sunda Shelf. The resulting RMS error in Figure 2 (marked by "R") is reduced by 30% from the control experiment in summer with little change in other months, suggesting large uncertainties in Wyrtki's transport estimates in summer.

5. Data assimilation with OI and PSAS

We demonstrate the use of the OI and PSAS schemes in the assimilation of high-frequency (HF) radar-derived surface velocity fields in the Monterey Bay area in California. The model is a fine-resolution, sigma coordinate version of Blumberg and Mellor (1987) hydrodynamic model. In the experiment, covariance matrices $\mathbf{P}^f$ and $\mathbf{R}$ are derived from estimates of horizontal covariance of the observed surface current data. It is found that the covariance functions are approximately Gaussian distributed. The use of a vertical projection method is not necessary, because subsurface velocity is less correlated with the surface currents.

Around the Monterey Bay area during the upwelling season, flow in the upper 70 meters of the water column is predominantly alongshore and southward with occasional directional fluctuations following the wind fields (Rosenfeld et al., 1994). Figure 3a shows the radar-derived surface current on August 15, 1995. A northward current appears inside the Monterey Bay, turns westward at 36.85°N, and forms a cyclonic eddy centered at 36.75°N, 122.2°W. This cyclonic eddy is not reproduced by the model without data assimilation (Figure 3b) probably because of coarse resolution of the wind forcing. The current fields are dominated by a southeastward velocity component except for an eastward current near the coast. When the surface velocity is assimilated into the model over a one-month period using the PSAS scheme, a cyclonic eddy centered at 36.75°N, 122.25°W is evident after 22 days of simulation (Figure 3c). The result is similar in the simulation using OI (figure not shown). Thus, the cyclonic eddy feature could be reproduced by assimilating observed surface currents.

6. Summary and Conclusions

Data assimilation produces model states that are representative of the observations and yet are constrained by physics in the numerical model. The mesoscale variability in sea level, velocity, and temperature fields are better resolved. Without data assimilation, the large RMS error between the simulated SSH and T/P SSH in the South China Sea indicates inadequate spatial resolution in the NCEP/NCAR winds and uncertainties in Wyrtki's transport estimates. With data assimilation, a reduced RMS error demonstrates that uncertainties associated with the forcing field and boundary conditions could be remedied to certain

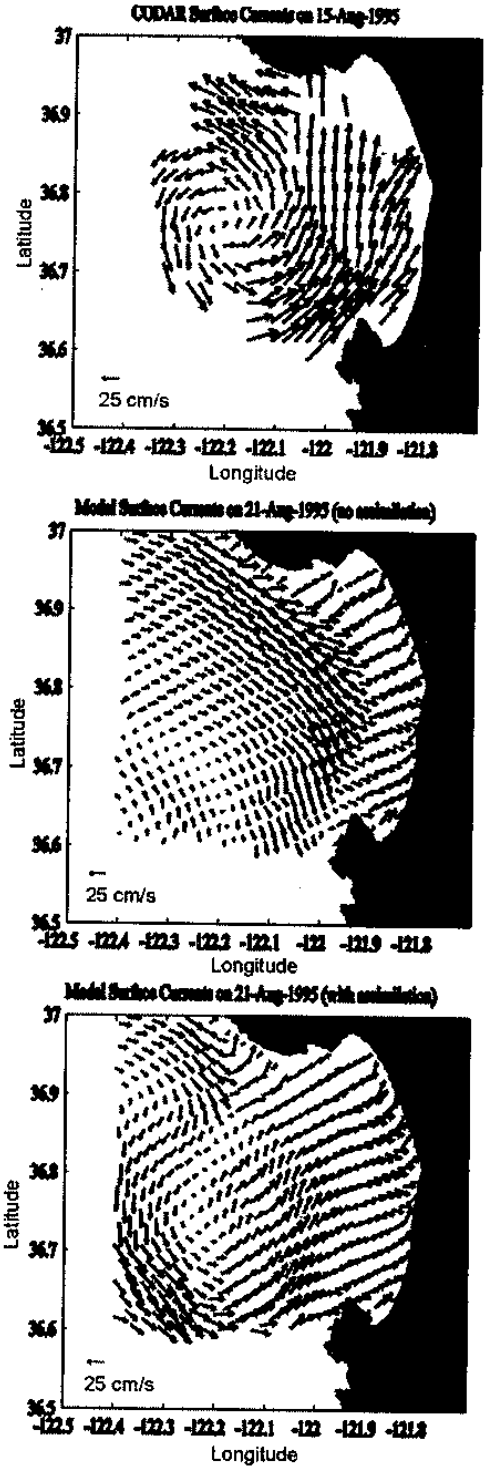

Figure 3. Sea surface currents in the Montery Bay area in California. (a) From radar observation. (b) model simulation without data assimilation, and (c) simulation with radar data assimilated. The velocity scale is 25 cm/s.

extent. The assimilation experiment using high-frequency radar-derived surface velocity fields in the Monterey Bay area in California shows significantly improvement in the model predictions of surface circulation. Thus, the OI and PSAS schemes have the potential of providing a better description of the velocity field in the South China Sea and better estimates of the transport through the Luzon Strait.

7. References

Anthes, D. L. T., Data assimilation and initialization of hurricane-predicting models. *J. Atmos. Sci.*, **31**, 702-719, 1974.

Blumberg, A., and G. L. Mellor, A description of a three-dimensional coastal ocean circulation model. In *Three Dimensional Coastal Models*, N.S. Heaps, Ed., Coastal and Estuarine Sciences, **4**, Am. Geophys. Un., 1-16, 1987.

Callahan, P. S., *TOPEX/POSEIDON GDR User's Handbook*, JPL D-8944, Jet Propulsion Laboratory, Pasadena, CA, 81 pp, 1994.

Chao, S.-Y., P.-T. Shaw and S. Y. Wu, El Nino modulation of the South China Sea circulation. *Progress in Oceanography*, **38**, 51-93, 1996a.

Chao, S.-Y., P.-T. Shaw and S. Wu, Deep water ventilation in the South China Sea. *Deep-Sea Res.*, Part I, **43**, 445-466, 1996b.

Cohn, S. E., A. da Silva, Jing Guo, M. Sienkiewicz, D. Lamich, Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Monthly Weather Review*, **126**, 11, 2913-2926, 1998.

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

Daley, R. *Atmospheric data analysis*. Cambridge University Press, NY, 457 pp, 1991.

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.

Haines, K., A direct method of assimilating sea surface height data into ocean models with adjustments to the deep circulation. *J. Phys. Oceanogr.*, **21**, 843-868, 1991.

Hurlburt, H. E., D. N. Fox, and E. J. Metzger, Statistical inference of weakly correlated subthermocline fields from satellite altimeter data. *J. Geophys. Res.*, **95**, 11375-11409, 1990.

Kalnay, E., M. et al., The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471, 1996.

Levitus, S., *Climatological atlas of the world ocean*. NOAA Professional paper No. 13, U. S. Government Printing Office, Washington, D. C., 173 pp, 1982.

Rosenfeld, L.K., F.B. Schwing, N. Garfield, and D.E. Tracy, Bifurcated flow from an upwelling center: A cold water source for Monterey Bay, *Continental Shelf Research*, **14**, 931-964, 1994.

Shaw, P. -T. and S. -Y. Chao, Surface circulation in the South China Sea. *Deep-Sea Res.*, part I, **41**, 1663-1683, 1994.

Shaw, P.-T., S.-Y. Chao, K.-K. Liu, S.-C. Pai and C.-T. Liu, Winter upwelling off Luzon in the northeastern South China Sea. *J. Geophys. Res.*, **101**, 16435-16448, 1996.

Shaw, P. -T., S. -Y. Chao and L. -L. Fu, Sea surface height variations in the South China Sea from satellite altimetry, *Oceanologica Acta*, **22**, 1-17, 1999.

Wu, C. -R., P. -T. Shaw, and S. -Y. Chao, Seasonal and interannual variations in the velocity field of the South China Sea. *J. Oceanogr.*, **54**, 361-372, 1998.

Wu, C. -R., P. -T. Shaw, and S. -Y. Chao, Assimilating altimetric data into a South China Sea model, *J. Geophys. Res.*, **104**, 29987-30005, 1999.

Wyrtki, K., Physical oceanography of the southeast Asian waters. NAGA Report Vol. 2, Scientific Results of Marine Investigation of the South China Sea and the Gulf of Thailand, Scripps Institution of Oceanography, La Jolla, California, 195 pp, 1961.