Implementation of Data Assimilation in ROMS model for Indian Ocean

Md. Nurujjaman
Department of Physics
National Institute of Technology Sikkim
Ravangla, Sikkim-737139, India
Email: jaman_nonlinear@yahoo.co.in

Amit Apte
Centre for Applicable Mathematics, TIFR
Sharada Nagar, Chikkabommasandra
Bangalore 560065, India
Email: apte@math.tifrbng.res.in

P. N. Vinayachandran
Centre for Atmospheric and Ocean Sciences
Indian Institute of Science, Bangalore, India,
Email: vinay@caos.iisc.ac.in

A. B. Samaddar
Department of Computer Science,
National Institute of Technology Sikkim
Ravangla, Sikkim-737139, India
Email: absamaddar@gmail.com

***Abstract*—Study of ocean dynamics and forecast is crucial as it influences the regional climate and other marine activities, and hence forecasting oceanographic states like sea surface temperature (SST), sea surface currents and mixed layer depth at different time scales is very important. These forecasts are generated by various ocean general circulation models. One such model is the Regional Ocean Modelling System. Though it is a complex model and reproduces many features of the ocean, it fails to reproduce the thermocline and hence one need to incorporates *Data Assimilation* (DA) in the model. DA system using ensemble transform Kalman filter has been developed for this model to improve the accuracy of the forecast. Temperature and salinity data from Argo float has been assimilated in the model. Updated SST shows that the sparsely populated observations of the oceanic states affects the states of the entire ocean, such effects get localized within the radius of influence of the observations by introducing the localization.**

***Index Terms*—Data Assimilation, Ensemble Transform Kalman Filter, ROMS, Indian Ocean, Oceanography, nonliearity, forecast**

I. Introduction

Oceans play a crucial role in regulating the regional climate, and other marine activities ranging from conventional fishing to high-tech oil and natural gas exploration. Tropical atmospheric convection is highly sensitive to the oceanic states like sea surface temperature (SST) that in turns affects the climate the region around the ocean [1], [2]. Rapidly growing economy of India is highly dependent on the climate and vulnerable to the the flood from the intense convective and coastal storms. More than ever, the Nation needs the benefit of more timely and accurate forecasting of the climate, which in turn highly dependent on oceanographic states.

Ocean General Circulation Models (OGCM) is customized for particular ocean to simulate and predict its features realistically to generate oceanic states forecasts. One such widely used OGCM for Indian Ocean is the *Regional Ocean Modeling System* (ROMS) [7]. Though ROMS is a complicated model, which uses parallel computation and many other parameterization scheme, is able to reproduce many oceanic states realistically. But some other important features like thermocline cannot be reproduced accurately. As ROMS is also nonlinear and does not capture all the physics of the ocean it produces erroneous forecast after certain time whatever accurate the initial conditions are used. Therefore, the models need to be re-initialised or adjusted with the observations. Incorporating observations effectively in prediction models is the subject of *data assimilation* (DA), and a good DA system is the key to improve ocean forecasting [3], [4], [5], [6]. In the broad sense, oceanic DA is the incorporation of the temperature, salinity and other oceanic state measurements into the OGCM. The goal of DA is to produce a regular, physically consistent forecast of the oceanic states from the model by using the irregularly sampled observations.

There have been many reports of successful implementation of DA in the various weather and ocean predicting centres, where research is going on various aspect of DA to improve the forecast of the OGCM, Numerical weather model and ocean-weather coupled model. Considering the impact of ocean and weather forecast on the economy of India, initiative towards improving the forecast by implementing DA is very much necessary. Also as a huge amount of data is available due to the successful operation of ARGO float, weather satellite and Radar, successful implementation of the DA system for Indian Ocean will be very useful to improve the forecast.

In this paper, we have presented the implementation and results of the simulation of ROMS model by using a DA system developed for the Indian Ocean, and the paper is organized as follows: Description of the model and DA scheme have been presented in Sec. II and III respectively. Results are presented in Sec. IV. Finally, conclusion is drawn in Sec. V.

II. Regional Ocean Modeling system (ROMS)

A. Equations of Motion

The basic features of the ROMS model [7], [8], [9], [10] are as follows: It uses primitive equations with potential temperature, salinity and an equation of state, where hydrostatic and

Boussinesq approximations are used for pressure. The model is solved horizontally using orthogonal-curvilinear coordinates on Arakawa C grid, whereas vertically terrine-following s-coordinate is used. Also along Horizontal direction Laplacian and biharmonic mixing options and vertically harmonic viscosity are used [11]. The primitive equations in Cartesian coordinates are given below.

The momentum balance in the x- and y-directions are:

$$\frac{\partial u}{\partial t}+\vec{v}\cdot\nabla u-fv=-\frac{\partial \phi}{\partial x}-\frac{\partial}{\partial z}\left(\overline{u'w'}-\nu\frac{\partial u}{\partial z}\right)+\mathcal{F}_u+\mathcal{D}_u \quad (1)$$

$$\frac{\partial v}{\partial t}+\vec{v}\cdot\nabla v+fu=-\frac{\partial \phi}{\partial y}-\frac{\partial}{\partial z}\left(\overline{v'w'}-\nu\frac{\partial v}{\partial z}\right)+\mathcal{F}_v+\mathcal{D}_v \quad (2)$$

Evolution of the salinity, temperature and nutrients $[C(x,y,z,t)]$ is governed by the advective-diffusive equation:

$$\frac{\partial C}{\partial t}+\vec{v}\cdot\nabla C=-\frac{\partial}{\partial z}\left(\overline{C'w'}-\nu_\theta\frac{\partial C}{\partial z}\right)+\mathcal{F}_C+\mathcal{D}_C \quad (3)$$

The equation of state is given by:

$$\rho=\rho(T,S,P) \quad (4)$$

$$\frac{\partial \phi}{\partial z}=-\frac{\rho g}{\rho_o} \quad (5)$$

The final equation expresses the continuity equation for an incompressible fluid:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \quad (6)$$

where $\mathcal{D}_u, \mathcal{D}_v$ and $\mathcal{D}_C$ are diffusive terms. $\mathcal{F}_u, \mathcal{F}_v, \mathcal{F}_C$ are forcing terms. $f(x,y)$ is Coriolis parameter; g is the acceleration due to gravity. $h(x,y)$ is the bottom depth; ν and ν_θ are molecular viscosity and diffusivity; P is the total pressure ($P \approx -\rho_o g z$). $\phi(x,y,z,t)$ is the dynamic pressure, where $\phi=(P/\rho_o)$. $\rho_o+\rho(x,y,z,t)$ is the total *in situ* density. $\zeta(x,y,t)$ is the surface elevation.

B. Variable position in Grid

The model equations are solved on a square domain $[30^0 S-30^0 N$ and $30^0 E-120^0 E]$. The variables positions are shown in horizontal ROMS grid is shown in Fig. 1. Temperature (T), salinity (S) and ocean dynamic-height (ζ), density (ρ), f and Ω are evaluated at ρ-points. u and v variables are evaluated at u and v-point respectively. Land regions are masked by defining on ρ-points. It is also defined on u, v and ψ-points.

C. Discretization of the variables

In the horizontal directions (ζ,η) second-order finite-difference approximation is used on the Arakawa "C" grid [12]. Placement of the variables in this grid is shown in Fig. 2. Same discretization method is also used in vertical direction. Evaluation steps of different variables are shown in the same figure.

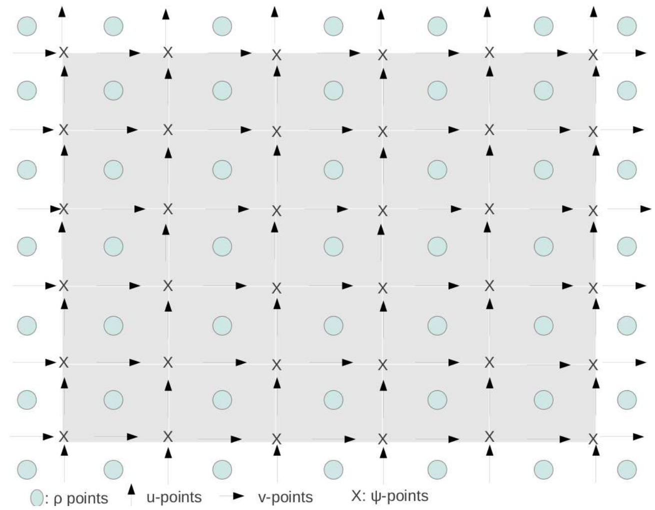

Fig. 1. position of the ocean variables on the grid.

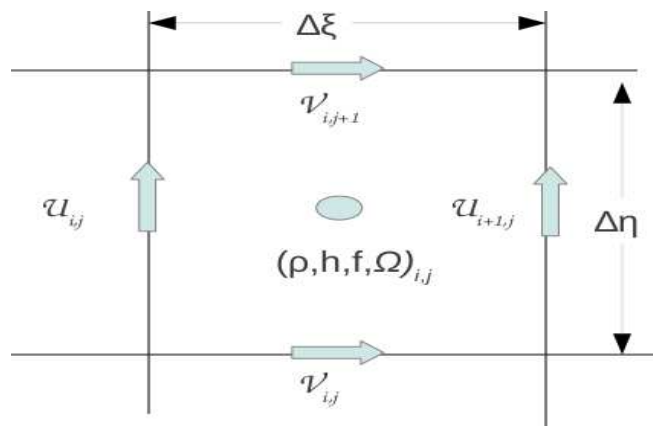

Fig. 2. Position and evaluation sequence of the variables on the Arakawa grid.

D. Evaluation of CPU time

The model works based on the parallel computation, where whole ocean domains is divided into several square tiles. All the tiles are evaluated simultaneously. We have estimated the CPU time for ROMS model for one day model run with different processor numbers, and model run time (hr) vs. number of processors plot is shown in Fig. 3. It shows that when number of processors are less, time taken to run the model for one day decreases with CPU numbers exponentially and becomes almost constant for higher number of processors.

III. Data assimilation (ETKF)

DA is a powerful and versatile method for combining partial, noisy observational data of a system with its dynamical model, generally numerically implemented, to generate state estimates of nonlinear, chaotic systems. Assimilation of the observations in the model improves the model state, and provides the most probable state of the system, given the

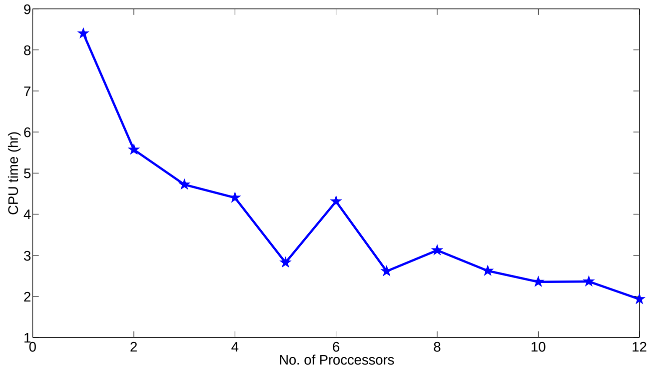

Fig. 3. CPU time taken by one day ROMS model run.

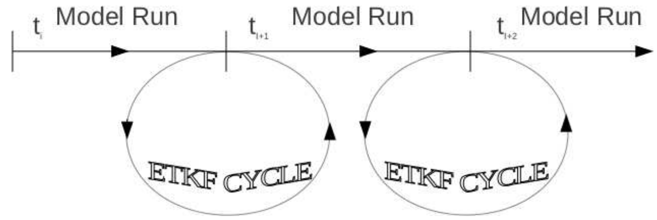

Fig. 4. Pictorial view of ROMS model run and update using ETKF

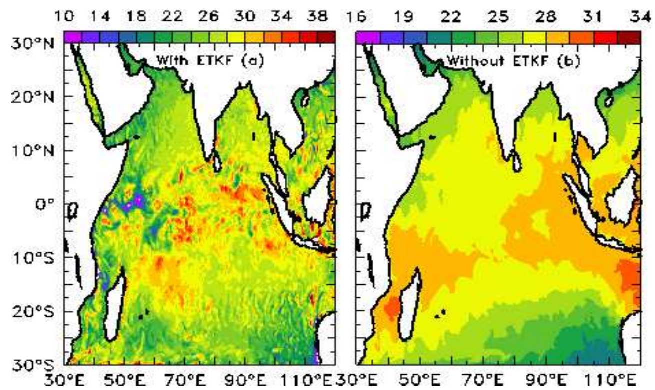

Fig. 5. First column shows the sea surface temperature (SST) of the update using ETKF, and the second panel shows one day model run without assimilation. Effect of DA through out the ocean is clear from first plot.

uncertainties in the observations and model forecast [13], [14], [15], [16].

DA produces a regular, physically consistent, initial conditions (x^a) for the ocean from a irregular collection of in situ temperature and salinity observations (y^o) from ARGO floats and model output (x^f). In this case x^a is a linear weighted mean between x^f and y^o based on their respective error covariances so that the linear combination of all the available information has the smallest possible total error covariance, and this linear combination can be obtained by Kalman Filter. The equations for Kalman filter are as follows:

$$X^a = X^f + K(y^o - HX^f) \quad (1)$$
$$P^a = (I - KH)P^f \quad (2)$$

where, $K = P^f H^T (HP^f H^T + R)^{-1}$ the Kalman gain.

Without going into the detail, one can derive the expressions for the forecast anomaly and updated mean respectively as

$$Z_e^a = \tilde{A} Z_e^f \quad (3)$$

$$\bar{X}_e^a = \bar{X}_e^f + K_e(\bar{y} - H\bar{X}_e^f). \quad (4)$$

for nonlinear and high dimensional system. Where, $Z_e^f = \{\delta x_1^f \quad \delta x_2^f \quad ... \ \delta x_N^f\}$ and $Z_e^a = \{\delta x_1^a \quad \delta x_2^a \quad ... \ \delta x_N^a\}$. $\tilde{A}$ is transformation matrix.

Eqn. 3 and 4 together forms the update for the ensemble transform Kalman filter (ETKF). Model and ETKF run has been shown graphically in Fig. 4. Ensembles of model run from some time (t_i) for one day to (t_{i+1}), and at this point forecast is updated using the available observations using ETKF. The analysis is used to run the model again upto the next update time (t_{i+2}) and so on.

Temperature (T), salinity (S), horizontal velocities (u and v) and dynamics ocean height (ζ) have been taken as prognostic variables of the dynamics, so they are used to form a state vector (a column vector). Each column of the Z_e^f matrix represents the column vector formulated by using the prognostic variables of different model runs simultaneously. Each column of Z_e^f represent the total number of model runs with different initial conditions, and hence they form an ensemble. In final Z_e^f ensemble mean ($\bar{x}$) is subtracted to perform the DA operations, where, $\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$, here the variable x represents any of the prognostic variables.

Similarly, the salinity and temperature from the ARGO data have been used to form observations vector (y^o).

IV. Results and Discussion

DA has been attracted wide spread use in oceanic model to improve the forecasting capabilities of the existing OGCM. Various DA schemes have been tested for such performance. Here we have presented assimilated SST and velocity fields using ETKF scheme.

Fig 5 shows the comparison of the one step update using ETKF. The updated SST is shown in Fig. 5(a) and model run without ETKF in Fig. 5(b). Fig. 5(a) shows the updated SST when approximately 2000 observations spread over the whole Indian Ocean were used. In this figure the fluctuations in the SST indicates that these sparsely populated observations affect the whole sea temperatures, which is due to the spurious correlation between observations and oceanic state. Such undesired correlations have been removed using localization scheme using Schur product [17], [18], and the detail of the scheme has been described in Ref. [17].

Fig. 6 shows the SST with localization with a scale of influence 200 Km. The figure in the first column shows the updated SST using sixty ensembles, second column shows updated SST using forty ensembles and their differences in the third column. The figures clearly show that the effect of the observations on the ocean temperature get localized. The patches show the updated SST, which increases at the

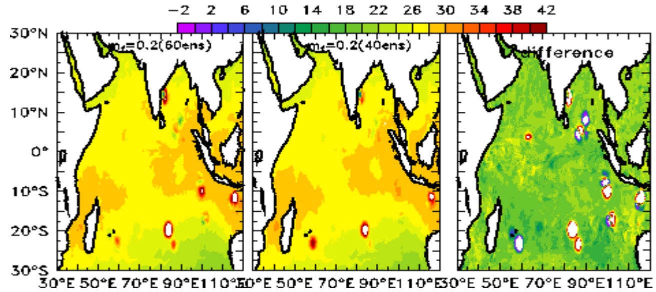

Fig. 6. Data assimilation using localization with localization scale 200 km. First column: sea surface temperature (SST) using sixty ensemble; second column: SST using forty ensemble and third column difference between first and second column.

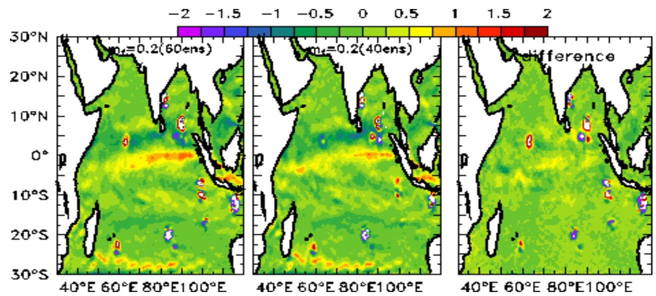

Fig. 7. Data assimilation using localization with localization scale 200 km. First column: u using sixty ensemble; second column: u using forty ensemble and third column difference between first and second column.

positions of the ARGO float. The patches also shows that the effect of observations are limited inside the influence circle of radius 200km. Inside the circle of influence temperature diverges from the normal temperature.

Almost same kind of behavior was observed in the velocity fields. Fig. 7 shows that u-velocity also changes considerably around the ARGO float within the radius of influence. Plot in Fig. 7 (first column) shows updated u-velocity using sixty ensembles, plot in the second column is the forty ensembles assimilation, whereas in the third column difference between the update shown in fist and second column are presented. The figures show that the temperature and salinity observations also affect the velocity fields in the same manner as of SST, and the effect get localized within the radius of influence around the observations.

As without localization assimilation affects whole ocean, which is due to the effect of spurious correlations, we have introduced localization in DA system. Both DA without and with localization is not giving ambient results so far. As the model run also diverged after few assimilation steps, a new parameter called *mean factor* (MF) was introduced to minimize the divergence of the updated mean [Eqn. 4]. On introduction of the MF, we are able to run the model with DA system for 17 days. However, there are several issues that we have to fixed are discussed in the following section.

V. Conclusion

We have developed a DA system using ETKF for Indian Ocean. With introduction of the localization, we are able to localize the effect of the DA around the observations. Though the entire system currently works for a month, there are several issues to get an improved DA system, and they are (a) the estimations the error covariance from model and observations, (b) designing effective initial perturbation fields for ROMS model, and (c) determination of the minimum ensemble size required for effective forecast.

Finally, the development of an effective DA system for Indian Ocean will improve the forecast of the oceanic sates at appropriate spatial and temporal resolutions, and contribute to Weather/Monsoon/Climate forecast by providing forcing for atmospheric models.

Acknowledgments

We acknowledge Sumanth Bhardwaj for his help to run various computational job for this works. Ferret has been used to analyze the data.

References

[1] P.N. Vinayachandran, D. Shankar, J. Kurian, F. Durand and SSC Shenoi, current Science **93**, 203 (2007).

[2] D. Shankar, P.N. Vinayachandran, A.S. Unnikrishnan, *The monsoon currents in the north Indian Ocean*, Progress in Oceanography **52**, 63−120 (2002).

[3] J. Whitaker and T.M. Hamill, *Ensemble data assimilation without perturbed observations*, Mon. Wea. Rev. **130**, 1913−1924 (2002).

[4] Pavel Sakov and Laurent Bertino, *Relation between two common localisation methods for the EnKF*, Comput. Geosci. (2010).

[5] JS, Hamill TM, Wei X, Song Y, Toth Z, *Ensemble data assimilation with the NCEP global forecast system*, Mon Wea Rev **136**, 463−482 (2008).

[6] *Modern Approaches to Data Assimilation in Ocean Modeling*, edited by P. Malanotte-Rizzoli (Elsevier Science, 1996).

[7] Alexander F. Shchepetkin, James C. McWilliams, *The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model*, Ocean Modelling **9**, 347−404 (2005).

[8] Katherine S Hedström, *Draft Techninal Manual for a Coupled Sea-Ice/Ocean Circulation Model*, 2000.

[9] Y. Song and D.B. Haidvogel, *A semi-implicit ocean circulation model using a generalized topography-following coordinate system*, J. Comp. Phys. **115(1)**, 228-244, 1994.

[10] D. B. Haidvogel and Beckmann, Numerical Ocean Circulation Modeling, Imperial College Press. 1999.

[11] R.C. Pacanowski and G. H. Philender, *Parameterization of vertical mixing in numerical models of tropical oceans*, J. Phys. Oceanography **11**, 1443-1451, 1981.

[12] A. Arakawa and V. R. Lamb, Method of computational physics **17**, 174-265, Academic Press, 1977.

[13] Eugenia Kalnay. *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, (2003).

[14] P. Malanotte-Rizzoli, editor. *Modern Approaches to Data Assimilation in Ocean Modeling*. Elsevier Science, (1996).

[15] G. Evensen. Springer, (2009).

[16] Kayo Ide and Christopher K.R.T. Jones. Nos. 1-2, *Data Assimilation (special issue) Physica D* **230**, (2007).

[17] R. Petrie. *Localization in the Ensemble Kalman Filter*. Masters thesis, Department of Meteorology. Reading University, August 2008.

[18] Anderson JL, *Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter*, Physica D **230**, 99111 (2007).