

Konstantin P. Beljaev (Center for Weather Forecast and Climate Studies, Brazil)  
 Clemente A.S. Tanajura (National Laboratory for Scientific Computation, Brazil)  
 James J.O'Brien (Florida State University, USA)

1. Introduction

This data assimilation method based on the well-known Kalman filter is implemented in combination with the Modular Ocean Model (MOM) of Geophysical Fluid Dynamics Laboratory (GFDL). This method is applied in combination along with the observational surface and subsurface temperatures from the PIRATA data set.

In this paper, a new method for the Kalman filter of the error covariance matrix is presented. It compares the studies published in literature by authors (Beljaev, K. S. Meyers and J. O'Brien, 1999). The present method is applied to the GFDL primitive equation ocean model MOM2 (Bjorn, 1969). The goal is to demonstrate the usefulness of the assimilation technique. This model is used only as a tool for the simulation of the thermodynamic fields, their prediction and variability. It has wide applications. Many papers discuss the model and its developments. Also, the Pilot Research Moored Array on the Tropical Atlantic (PIRATA) data set is used here. This includes sea surface and subsurface temperatures records.

2. Data Assimilation Technique.

The following problem is considered. Let  $\zeta(t, \bar{x})$  be an unknown real which is sought in some ocean domain  $\Omega$ . The model equation for  $\zeta$  is supposed to be written in Cartesian coordinate system as below:

$$\frac{\partial \zeta}{\partial t} = A(t)\zeta,$$

where  $A(t)$  is the known model operator, in general non-linear.

Let us define the model solution  $\zeta(t, \bar{x})$  up to the time  $t_0$  and to the coordinates  $(\bar{x})$  in the domain  $\Omega$ . Let  $\tau_0$  denote the time of the first observation. Let  $\bar{x}_i$  denote the coordinates of the  $i$ -th observation point. Let  $\tau_i$  denote the time of the  $i$ -th observation. Let  $\tau$  denote the time of the  $N$ -th observation. The problem is considered the opposite to the usual one: only the observations are known and the model is unknown.

Let  $\hat{\theta} = \theta(t, \bar{x}) = \zeta(t, \bar{x}) - \zeta_{\text{mod}}(t, \bar{x})$  be an error in modeling or simply error. The exact solution is to be satisfied for  $\theta$ :

$$\frac{\partial \theta}{\partial t} = A(t)\theta + H$$

And also, the standard conditions for the zero:

$$L\eta = 0, L\eta(\bar{x}, \tau)\eta(\bar{y}, t) = R(t)\delta(t - \tau)$$

are assumed to hold. All notations are non vector. The function  $R$  is known.

The following problem is considered: to find out the optimal estimation  $\hat{\zeta}(t, \bar{x})$  of  $\zeta(t, \bar{x})$  after the  $N$  unknown observations  $\zeta(\tau_i, \bar{x}_i)$  satisfying the conditions:

a)  $\hat{E}(\zeta - \hat{\zeta}) = 0$

b)  $\hat{E}(\hat{\zeta} - \zeta)^2 = \min_{\hat{\zeta}} E(\zeta - \hat{\zeta})^2$  in all estimation.

The optimal filter is sought as follows:

$$\hat{\zeta}(t, \bar{x}) = \zeta_{\text{mod}}(t, \bar{x}) + \int_0^{N(t)} \sum_{i=0}^{N(t)} \alpha(\tau, \bar{x}, \bar{x}_i) \theta(\tau, \bar{x}_i) d\tau$$

Let  $N(\tau)$  be the number of observations up to time  $\tau$ . In formula (6), the unknown weight-coefficients  $\alpha_i = \alpha(\tau, \bar{x}, \bar{x}_i)$  should be determined using condition b).

The following equation for  $\alpha$  is

$$K(t, \bar{x}, \bar{x}_i) = \int_0^{N(\tau)} \sum_{j=0}^{N(\tau)} \alpha(\tau, \bar{x}, \bar{x}_j) K(t - \tau, \bar{x}_i, \bar{x}_j) d\tau$$

where  $K = K(t, \bar{x}, \bar{y})$  is the error covariance function. Another way is suggested to define the covariance function and its time evolution.

The probability distribution is found from the Euler-Picard eq.

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial x} + \frac{\partial p}{\partial y} + \frac{1}{2} (B \frac{\partial^2}{\partial x^2}) p = 0$$

where the operators are defined both from the model and observations. Then the covariance found, by gradually knowing the probability distribution and its time evolution.

3. The numerical experiments and result

The results of the data assimilation are analyzed using the PIRATA data. Some general conclusions are made. Also, the first series of experiments with the model is compared with the observations. The results are presented in the next section.

starting from the known initial temperature-salinity field described above. The real data are assimilated on the same day. Model predicted fields along with observations create the new ocean state which is taken as an initial condition for the next time step (next day). This continues during one month. For the next month the model starts again from the climate initial state.

Let  $\sigma_a^2(t)$  be the model error variance without any assimilation, i.e. the error between the model values themselves and observations. Along with the variable  $\sigma_a^2(t)$  two other variables are considered:  $\sigma_b^2(t)$  and  $\sigma_c^2(t)$ . They are the variances before and after correction, respectively. The first variable shows the difference between the model values and observations at moment  $t$ , if the correction is not done at the previous time step, and  $\sigma_c^2(t)$  shows the difference after the correction at the same time step.

Fig. 1 illustrates the time behaviour of these three variables through March '99 at a point located at the 10-meter depth. The top curve shows the model error variance  $\sigma_a^2(t)$ . It remains on first steady during the month (around 1.5 to 2 square degrees). A "spike" is observed in the end of the month due to the appearance of a warm anomaly near the equator. The model could certainly not predict or react to it. Therefore, this spike was inevitable. The second curve demonstrates the time-behaviour of variance  $\sigma_b^2(t)$ . One can see that in the first four days this value is even greater than the first error. It means that the model needs an "adjustment time" to come to an agreement with observations. But after four days this curve "plunges" under the first one and its values remain significantly smaller during all other days. Also, in the end of the month a similar spike is observed, but the method reacts quickly and precisely. It forces the model to hold the same level of error (around 0.5 square degrees) as at the beginning of the curve, i.e.  $\sigma_a^2(t)$ . Its values are also constant around 0.25 square degrees long through this period. This confirms that the method works correctly and its results are more like observations.

This method has been compared with a conventional version of the Kalman-filter.

Fig. 2 shows the comparison between the two methods. The results were taken at the 10-m level in January '99. As in Fig. 1, the top curve shows the model error  $\sigma_a^2(t)$ . It provides the same observations. The model error grows during the month and it starts from the beginning of the

dash-dot line is the error variance before assimilation of the Kalman filter version and the last curve with "circles" is the error variance of the studied method. One can see that the presented method is better. It gives not only the smallest variance, but also a quicker adjustment after unexpected spikes (in the beginning of the month and around day 20). These jumps are inevitable because the model does not have a forecast block. It is forced only by climatological winds and heat-fluxes, and it is unable to predict the real variability in the ocean. Again it should be pointed out that both assimilation methods work correctly. They both give the model adjustment to observations.

**References**

Bejaoui K., S.Meyers, and J.O'Brien: 1999, Ekker-Planck equation application to ocean assimilation based on Kalman-filter theory. *J. of Mech. Sciences*, (in press).  
 Bryan K.: 1969, A numerical method for the study of the circulation of the World Ocean. *J. Comp. Physics*, 4, 347-370.

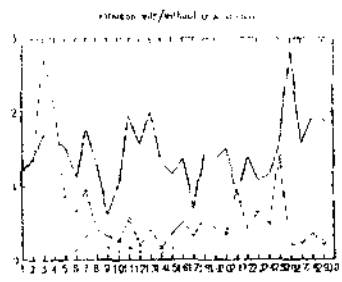


Fig.1. Variances with and without assimilation.

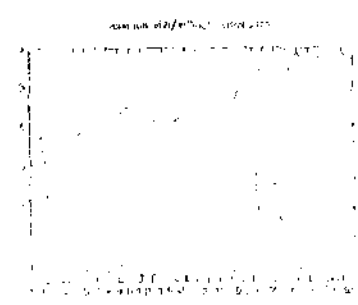


Fig.2. Comparison of the variances with standard Kalman-filter.