Ocean Modelling 116 (2017) 159–172

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

A balanced Kalman filter ocean data assimilation system with application to the South Australian Sea

Yi Li[a,b,*], Ralf Toumi[a]

[a] Space and Atmospheric Physics Group, Department of Physics, Imperial College, London
[b] National Marine Environmental Forecasting Center, State Oceanic Administration, China

ARTICLE INFO

Article history:
Received 18 October 2016
Revised 13 June 2017
Accepted 19 June 2017
Available online 20 June 2017

Keywords:
ROMS
EnKF
Multivariate balance

ABSTRACT

In this paper, an Ensemble Kalman Filter (EnKF) based regional ocean data assimilation system has been developed and applied to the South Australian Sea. This system consists of the data assimilation algorithm provided by the NCAR Data Assimilation Research Testbed (DART) and the Regional Ocean Modelling System (ROMS). We describe the first implementation of the physical balance operator (temperature-salinity, hydrostatic and geostrophic balance) to DART, to reduce the spurious waves which may be introduced during the data assimilation process. The effect of the balance operator is validated in both an idealised shallow water model and the ROMS model real case study. In the shallow water model, the geostrophic balance operator eliminates spurious ageostrophic waves and produces a better sea surface height (SSH) and velocity analysis and forecast. Its impact increases as the sea surface height and wind stress increase. In the real case, satellite-observed sea surface temperature (SST) and SSH are assimilated in the South Australian Sea with 50 ensembles using the Ensemble Adjustment Kalman Filter (EAKF). Assimilating SSH and SST enhances the estimation of SSH and SST in the entire domain, respectively. Assimilation with the balance operator produces a more realistic simulation of surface currents and subsurface temperature profile. The best improvement is obtained when only SSH is assimilated with the balance operator. A case study with a storm suggests that the benefit of the balance operator is of particular importance under high wind stress conditions. Implementing the balance operator could be a general benefit to ocean data assimilation systems.

1. Introduction

By using the Bayesian theorem, data assimilation provides an objective criterion for fusing observations with numerical models to produce an estimate of the true state (e.g. Lahoz and Schneider, 2014; Wikle and Berliner, 2007). This is a crucial step in providing an optimal initial condition for ocean forecasting (e.g. Lahoz and Schneider, 2014; Blayo et al., 2014). Global ocean data assimilation has developed rapidly during the past decade, and an increasing number of products are provided by groups worldwide (e.g., Global Ocean Data Assimilation Experiment, https://www.godae-oceanview.org/). However, the typical resolution of global models is still too low to resolve mesoscale features thus there is a need to develop regional ocean data assimilation systems (Moore et al., 2013).

There are two popular data assimilation approaches: variational and sequential (Lahoz and Schneider, 2014). The variational approach adjusts the model trajectories to fit the observations by minimising the cost function. On the other hand, the sequential approach updates the model state by comparing the mean and variability of both model and data, each time when the new observation is available. These two approaches are both based on the Bayesian theorem and can be simplified to similar algorithms under the assumptions of Gaussianity and linearity. The detailed description of data assimilation algorithms has been explored extensively (e.g. Lahoz and Schneider, 2014; Wikle and Berliner, 2007; Evensen, 2009).

The Kalman Filter (Kalman, 1960) is a sequential data assimilation algorithm designed for a linear dynamical model with Gaussian-distributed model and observation errors. Evensen (1994) developed the ensemble method by using a Monte Carlo technique to approximate the mean and covariance of a high-dimensional system. Compared with the variational data assimilation such as four-dimensional variational method (4DVAR), EnKF needs much less effort to implement since the tangent linear and adjoint models are not needed. It is also possible to use different physical schemes in the ensemble members.

EnKF has been used in several regional ocean data assimilation studies. Evensen and van Leeuwen (1996) developed the first EnKF

* Corresponding author.
E-mail addresses: y.li14@imperial.ac.uk, yli.ouc@gmail.com (Y. Li).

http://dx.doi.org/10.1016/j.ocemod.2017.06.007
1463-5003/

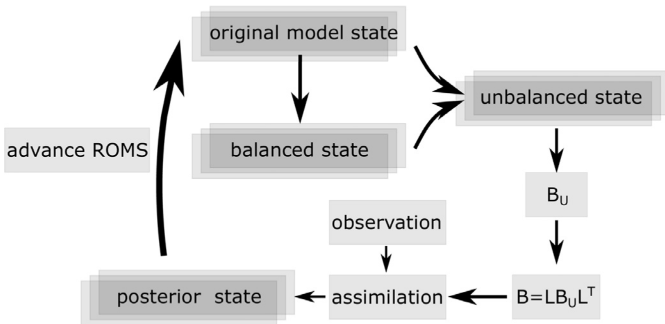

Fig. 1. A schematic description of the balanced data assimilation system.

based regional data assimilation system for the Agulhas current using a quasigeostrophic model. Seo et al. (2010) assimilated SST in a northwest Pacific ROMS model through a stochastic EnKF scheme, but only SST was assimilated and physical balance was not studied. Here we present the first implementation of DART (Anderson et al., 2009), a community data assimilation facility to ROMS, to provide several state-of-the-art algorithms with localisation, inflation techniques. Also, we use balance constraints to reduce the spurious waves generated in the data assimilation process.

The localisation technique is needed to reduce the harm caused by the spurious covariances with distant grid points in EnKF with limited ensemble members (Lorenc, 2003; Houtekamer and Mitchell, 2001; Gaspari and Cohn, 1999). However, localisation may cause imbalance to the dynamical system (Lorenc, 2003; Greybush et al., 2011), although a balanced model error covariance produces balanced analysis state (Cohn and Parrish, 1991). Lorenc (2003) pointed out that when SSH observation at a single point is assimilated without localisation, the increments of SSH and ocean currents are balanced. If localisation is used, the gradient of SSH increases while the ocean currents reduce. Therefore spurious ageostrophic waves are created in this process. A schematic description of this example can be found in Greybush et al. (2011) (in their Fig. 1). Mitchell et al. (2002) reported that different ensemble size, assimilating frequency or localisation radius can cause imbalance for GCMs. Kepert (2009) suggested using localisation of streamfunction and velocity potential instead of u and v velocities, but this method is not easy to implement for regional ocean models. For the multivariate problem, it is also difficult to specify the relationship among various variables. A common practice is to use linear regression, but this can be a major source of error (Anderson, 2007b). To solve this problem, Anderson (2007b) proposed to 'localise' the impact of observation to model states (e.g., SST observation at one point and simulated ocean currents at another point), but he also pointed out that it is usually difficult to define the 'distance' between them, especially for high-dimensional GCMs. In this paper, we use a multivariate balance operator proposed by Weaver et al. (2005) to solve the imbalance problem. In this algorithm, each model variable is separated into balanced and unbalanced components, and several balance assumptions are made to calculate the increments. This physical constraint has been used in several variational ocean data assimilation systems (e.g. Balmaseda et al., 2013; Li et al., 2008; Moore et al., 2011c).

The DART/ROMS data assimilation system is applied to the South Australian Sea (31.5°S–39.5°S, 117°E–140°E). This region hosts the world's longest zonal, mid-latitude shelf (about 2500 km) between Cape Leeuwin and Portland (Middleton and Bye, 2007). The Leeuwin Current, flowing southward from the tropics near the west coast of Australia, enters the South Australian Sea around Cape Leeuwin and extends to Tasmania. This region is recognised as one of the 64 Large Marine Ecosystems (LMEs) by NOAA (http://www.lme.noaa.gov/). There are also emerging tourism ventures and oil/gas exploration. The simulation and prediction of the ocean circulation, temperature and other oceanic variables are therefore necessary.

In this paper, we first describe the data assimilation system and the balance operator in Section 2. In Section 3 we evaluate the effect of balance operator using an idealised two-dimensional shallow water model. The results of the South Australian Sea model are given in Section 4. In Section 5 we discuss and analyse the results from Section 3 and 4. A summary concludes this paper in Section 6.

2. Method and data

The Ensemble Adjustment Kalman Filter (EAKF) (Anderson, 2001) is implemented as the data assimilation algorithm. We introduce a physical balance operator to this system and analyse the performance in both an idealised shallow water model and the ROMS real case.

2.1. Ensemble Kalman Filter and the balance operator

There are many implementations of the ensemble Kalman Filter (e.g. Evensen, 2003; Houtekamer and Mitchell, 2001; Anderson, 2001). The data assimilation cycle consists of two stages. Firstly, in the forecasting stage, the model state x evolves through a dynamic model and secondly in the analysis stage the estimation of model state is improved by comparing the forecast x^f and the observation y^o, the analysis is computed as,

$$x^a = x^f + K[y^o - H(x^f)], \quad (1)$$

where B is the model error covariance matrix, R is the observation error covariance. H is the observation operator that projects the model state x to the observation space $y = H(x)$. The difference between x^a and x^f is defined as the increment Δx. The optimal variance minimising weight is given by the Kalman gain,

$$K = BH^T(HBH^T + R)^{-1}, \quad (2)$$

In the ensemble data assimilation approach, the B matrix is computed from the N ensemble members in each data assimilation cycle as,

$$B = X^f X^{fT}, \quad (3)$$

where $X = \frac{1}{\sqrt{N-1}}(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_N - \bar{x})$ is the perturbation matrix whose columns are the deviations from the ensemble mean.

For an ocean model such as ROMS, there are 5 components in the state vector x: sea surface height η; potential temperature T; salinity S; horizontal velocities u and v. Temperature is usually the most observed variable in the ocean so Weaver et al. (2005) proposed to compute the relations between the variables based on T. Each variable except T is decomposed into two components, the balanced component and the unbalanced one. Therefore the perturbations δx can also be decomposed as following,

$$\begin{pmatrix} T \\ S \\ \eta \\ u \\ v \end{pmatrix} = \begin{pmatrix} T \\ S_B \\ \eta_B \\ u_B \\ v_B \end{pmatrix} + \begin{pmatrix} 0 \\ S_U \\ \eta_U \\ u_U \\ v_U \end{pmatrix} = L \begin{pmatrix} T \\ S_U \\ \eta_U \\ u_U \\ v_U \end{pmatrix}, \quad (4)$$

where the variables with subscript B represent the balanced component of the variables while those with subscript U represent the unbalanced one. The balanced part of variable x_1 can be derived from other variable x_2 through the linear balance operator L. The details of L are explained in Appendix A.

The model error covariance is thus converted to,

$$B = LB_uL^T, \quad (5)$$