Ocean Modelling 27 (2009) 160–173

Contents lists available at ScienceDirect

Ocean Modelling

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

Efficient sensor placement for ocean measurements using low-dimensional concepts

B. Yildirim [a], C. Chryssostomidis [b], G.E. Karniadakis [b,c,*]

[a] Brown University, Division of Engineering, Box D, Providence, RI 02912, USA
[b] Sea Grant College Program, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Brown University, Division of Applied Mathematics, 182 George St., Providence, RI 02912, USA

ARTICLE INFO

Article history:
Received 8 April 2008
Received in revised form 28 December 2008
Accepted 6 January 2009
Available online 14 January 2009

Keywords:
Data assimilation
Adaptive sampling
Proper orthogonal decomposition

ABSTRACT

Using simulation results from three different regional ocean models (HOPS, ROMS and FVCOM) we show that only a few spatio-temporal POD (proper orthogonal decomposition) modes are sufficient to describe the most energetic ocean dynamics. In particular, we demonstrate that the *simulated* ocean dynamics in New Jersey coast, Massachusetts Bay and Gulf of Maine is energetically equivalent to the wake dynamics behind a cylinder at low Reynolds number. Moreover, the extrema of the POD spatial modes are very good locations for sensor placement and accurate field reconstruction. We employ a modified POD theory to incorporate a limited number of measurements in reconstructing the velocity and temperature fields, and we study systematically the corresponding reconstruction errors as a function of the sensor location, number of sensors, and number of POD modes. This new approach is quite accurate in *short-term* simulation, and hence it has the potential of accelerating the use of real-time adaptive sampling in data assimilation for ocean forecasting.

1. Introduction

In order to initialize dynamical forecast systems for the ocean state, measurements of high accuracy are required; such measurements are difficult and costly, and in many cases, e.g., in real-time adaptive sampling, they have to be done very fast. Given the spatio-temporal variability in the ocean and its intermittent dynamics, sampling a pre-determined region *uniformly* in time–space can be very inefficient. This is due to the fact that only a small subset of those measurements have a significant effect on the accuracy of the forecasts (Robinson and Glenn, 1999). Adaptive sampling is an *evolving* method for the efficient sampling of the most energetic ocean phenomena in support of real-time nowcasting and forecasting. It has been used only recently in ocean forecasting demonstrations (e.g., HOPPS-ESSE, 2003) and has the potential of reducing the observational requirements by orders of magnitude. However, adaptive sampling is still complicated and costly for routine observations. Moreover, *truly* real-time adaptive sampling of the ocean requires *fast* data assimilation methods and *rigorous* criteria to identify locations of best sensor placement. To this end, adjoint methods (that solve the inverse problem) (Arango et al., 2003) bear great promise and a recent demonstration in a data assimilation experiment of the East Australia Current confirmed this (Wilkin et al., 2008). However, adjoint methods tax computational resources more heavily than other methods and they typically require a lot of computer memory. Other methods, e.g., based on uncertainty estimation such as the ESSE system (Lermusiaux, 2001), have been used in certain demonstration experiments but they too are expensive for *truly* real-time adaptive sampling. Various nonlinear versions of the Kalman filter have been applied to simplified models but no clear consensus on their effectiveness has been reached yet (Buehner and Malanotte-Rizzoli, 2003, Zang and Malanotte-Rizzoli, 2003). A recent progress report from a joint NSF-ONR workshop (Lermusiaux et al., 2006) on Data Assimilation (DA) identified the need for improving real-time adaptive sampling and recommended the development of new *economic* DA without loss of accuracy based on reduced-dimension schemes that will complement adjoint- and ensemble-based methods.

In light of ocean complexities over a wide range of scales – see the "multiscale ocean" in Dickey (2003) – extracting the proper hierarchy can be both valuable in physical understanding but also in developing new ways of modeling and forecasting ocean processes. Proper Orthogonal Decomposition (POD) (Rempfer, 2003; Bekooz et al., 1993; Aubry et al., 1991; Sirovich, 1987) is one such approach (also known as the method of Empirical Functions or EOF), and oceanographers have used it to analyze their data or to develop reconstruction procedures for gappy data sets (see Beckers and Rixen, 2003; D'Andrea and Vautard, 2001; Hendricks et al., 1996; Everson et al., 1995) and Wilkin and Zhang, 2006; Pedder and Gomis, 1998; Houseago-Stokes, 2000; Preisendorfer and Mobley, 1988). In the present work, we are interested in extracting

* Corresponding author. Tel.: +1 401 863 1217.
E-mail address: gk@dam.brown.edu (G.E. Karniadakis).

1463-5003/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2009.01.001

useful concepts from POD analysis that will lead to concrete guidelines for efficient sensor placement in adaptive sampling for regional ocean forecasting. Specifically, we pose the question: *Can we use properties of the POD modes, computed over a specific time interval, to decide on the best and hopefully optimum location of sensors?* Here we define optimum location as the positions that will give us the best possible reconstructed ocean field (in the mean energy sense), given a limited number of measurements at these locations.

To this end, instead of tracking instantaneous special flow features in the physical domain, we are looking for special features in the POD modes, that is in the *modal* domain. However, for this approach to be efficient, we first need to demonstrate, from numerical simulations, that the ocean dynamics predicted by different regional models is low-dimensional, and hence only a few dominant POD modes can reproduce the essential ocean dynamics. Clearly, the dimensionality of the ocean region we simulate depends strongly on the phenomena involved (e.g., convection, upwelling, etc.) and no general statements can be made. To address this difficult question, we have undertaken a systematic study using three different ocean models (ROMS, xxxx), (HOPS, xxxx) and (FVCOM, xxxx), to simulate the *short*- and *long-term* dynamics of different regions and phenomena. As we will see, this distinction between short- and long-term dynamics is important as the *effective dimensionality* of the ocean system increases with time. The above models employ assumptions associated with turbulence modeling and do not have the resolution fidelity to capture the small flow scales. In order to appreciate this, we also compare the POD analysis of the simulated ocean with a similar analysis of a benchmark problem, flow past a circular cylinder at low Reynolds number, which is based on outputs from high-resolution direct numerical simulation (DNS) with all spatio-temporal scales fully captured.

We present the simulation results in the next section. Subsequently, we present in Section 3 a recent extension of POD (Everson and Sirovich, 1995; Beckers and Rixen, 2003; Venturi and Karniadakis, 2004; Willcox, 2005), that leads to a reformulation of the data assimilation problem as a gappy data problem. First, we apply this approach to our benchmark problem, and then in Section 3 to Massachusetts Bay. We conclude the paper with a summary and a discussion of our findings in Section 4.

2. Simulations and low-dimensionality

We have employed three different regional ocean models, specifically, the Rutgers model (ROMS, xxxx), the Harvard model (HOPS, xxxx), and the University of Massachusetts at Dartmouth model (FVCOM, xxxx) to obtain simulation data for different regions and conditions (New Jersey coast, Massachusetts Bay (Mass. Bay), and Gulf of Maine, respectively). We have analyzed the ocean dynamics over short-term but also for much longer periods using POD. By cross-correlating different snapshots obtained from the simulations we constructed the covariance matrix, the eigendecomposition of which yields the POD eigenvalues and corresponding POD temporal and spatial modes. In particular, the sum of the normalized eigenvalues is representative of the energy captured by the corresponding POD modes. The POD modes are hierarchical with the lower-indexed modes containing higher energy. We used a serial code based on LAPACK for the ROMS and HOPS simulation outputs, however, we implemented a parallel version of the POD code based on ScaLAPACK to deal with the large matrices involved in the Gulf of Maine FVCOM simulations (matrices with more than 1 billion entries).

In the following, we describe the different data sets we analyzed and present a summary of our results that provides evidence of the low-dimensionality of the "numerical" ocean.

2.1. New Jersey and Middle Atlantic Bight

The Lagrangian Transport and Transformation Experiment (LaTTE) is a coordinated program of field and numerical experiments that addresses the biological and geographic extent of contaminants along the New Jersey and Middle Atlantic Bight (see LATTE (xxxx, 2007) and Fig. 1). We will use the LaTTE data base in our analysis. Here, POD analysis is carried out for two cases, one short-term and one long-term simulation. The simulation results were provided by the Rutgers Ocean Modeling Group (Rutgers Group, xxxx). The short-term simulation is for 2.5 days; it starts at midnight of May 13, 2005 and ends at noon of May 15, 2005. The long-term simulation is for 25 days; it starts on February 4, 2006 and ends on February 28, 2006. Time resolution of snapshots and various snapshot length effects on energy spectra and POD coefficients were investigated systematically to avoid any erroneous conclusions.

The normalized energy spectra for both cases are shown in Fig. 2 for three variables, namely horizontal velocity vector, temperature and salinity. We note that POD analysis was performed on the horizontal velocity *vector* and not on the components so only one line representing the velocity is shown in the plot. Moreover, the entire computational domain is employed in the POD analysis. Several resolution checks were performed to assess the accuracy of our results. In Fig. 3 we show the effect of sampling at four different resolutions. Specifically, the short-term simulation was sampled every 0.625, 1.25, 2.50 and 3.75 h while the long-term simulation every 3.0, 6.0, 12.0 and 18.0 h. The percentage of total energy ($\sum \lambda \times 100$) captured by the coarse resolutions superimposed on the finest resolution shows that the low modes of the finest simulation comprise about 99% of the total energy.

2.2. Massachusetts Bay

The Massachusetts Bay (Mass. Bay) simulation data base was provided by the Harvard Ocean Prediction System Group (Lermusiaux, 2001), see Fig. 4. The short-term simulation covers 8 days, starting on August 25 and ending on September 2, 1998; the data are recorded at every hour (193 snapshots). The long-term simula-

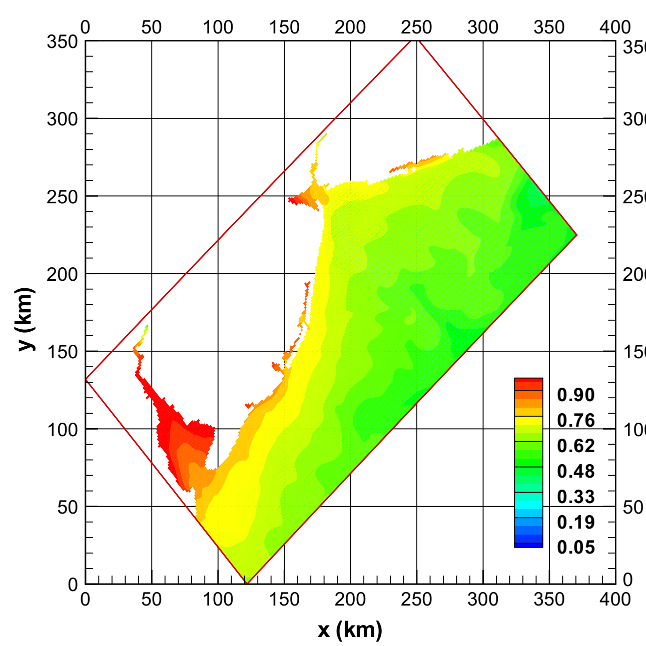

Fig. 1. *New Jersey coast:* Contours of the first POD mode (dimensionless temperature) and domain.