Efficient sensor placement for ocean measurements using low-dimensional concepts

Citation: Yildirim, B., C. Chryssostomidis, and G.E. Karniadakis. "Efficient sensor placement for ocean measurements using low-dimensional concepts." Ocean Modelling 27.3-4 (2009): 160-173. Web.

As Published: http://dx.doi.org/10.1016/j.ocemod.2009.01.001

Publisher: Elsevier Ltd.

Persistent URL: http://hdl.handle.net/1721.1/49867

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

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*

[b] *Sea Grant College Program, Massachusetts Institute of Technology, Cambridge, MA 02139*

[c] *Brown University, Division of Applied Mathematics, 182 George St., Providence, RI 02912*

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.

* Corresponding author
Email address: gk@dam.brown.edu (G.E. Karniadakis).

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 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 (URL)), (HOPS (URL)) and (FVCOM (URL)), to simulate the *short-term* and *long-term* dynamics of different regions and phenomena. As we will see, this distinction between short-term 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 (URL)), the Harvard model (HOPS (URL)), and the Uni-

versity of Massachusetts at Dartmouth model (FVCOM (URL)) 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 eigen-decomposition 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 (URL); Choi and Wilkin (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 (URL)). 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 0.625hr, 1.25hr, 2.50hr and

3.75hr while the long-term simulation every 3.0hr, 6.0hr, 12.0hr and 18.0hr. 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 eight days, starting on August 25 and ending on September 2, 1998; the data are recorded at every hour (193 snapshots). The long-term simulation covers 47 days, between August 20 to October 6, 1998; the data are recorded daily (47 snapshots). The energy spectra (excluding the mean mode) is given for both simulations in Fig. 5. As also seen in the New Jersey simulations, here too a few modes contain most of the energy for both simulations, see also Tables 1 and 2. The first four modes account for 93% for the 8-days simulation, and for 78% for the 47-days simulation, as seen in Table 1. Additionally, Table 2 shows the eigenmodal energy content of the instantaneous field (excluding the *mean* mode). The temperature and salinity variables show very low dimensionality for both simulations, with the first mode accounting for about 99% of the total energy.

2.3 Gulf of Maine

The Gulf of Maine was simulated by the Unstructured Finite Volume Coastal Ocean Model (USG-FVCOM) and the simulation data were provided by the University of Massachusetts-Dartmouth Ocean Modeling Group; see Fig. 6. The simulation covers all of August 1998 and the output is recorded hourly (744 snapshots); a typical result is shown in Fig. 7, including sensitivity of the velocity energy spectrum on two different grids. We note parenthetically here that the third POD mode shows that most of the activity is centered around Georges Bank (not shown here). Table 1 presents the first nine POD eigenvalues. The first mode of temperature and salinity again account for 99% of the total energy, while the sum of the first three eigenmodes of velocity accounts for about 95%.

2.4 The "numerical" ocean is low-dimensional

The physical and biological processes in the ocean are characterized by a wide range of spatio-temporal scales, from 1 mm for molecular processes to more

than 10 km for fronts, eddies and filaments, and corresponding characteristic times from 1 second to several months (Dickey (2003)). However, the question that we address here is what range of such scales is captured *in simulations* using some representative regional ocean models, and what is the corresponding energy hierarchy. (We refer to the *outputs* of the simulation codes as "numerical" ocean.)

We present a summary of all our results in terms of the relative energy as represented by the normalized eigenvalues. Specifically, in Table 1 we list the first nine POD eigenvalues from the analysis of the *total* fields, i.e., including the mean mode. This mode, corresponding to the first row in the table, is responsible for about 99.9% of the total energy for temperature and salinity. Velocity, however, has a wider distribution with its first two eigenmodes containing about 90% energy for the Gulf of Maine, about 70% for Massachusetts Bay, and about 50% for the New Jersey Coast. The energy contained in the sum of the first eight eigenmodes comprises about 97% for the *short-term* simulation and about 91% for the *long-term* simulation. We also present the POD eigenvalues in Table 2 with a different normalization, i.e., we do not include the first (mean) mode, and hence it is easier to appreciate the energy content in all the modes representing spatio-temporal fluctuations.

In order to put these results into the proper context we compare next with a well-studied prototype flow, namely the laminar and turbulent wake. In past work, in a series of papers (Dong et al. (2006); Ma and Karniadakis (2002); Ma et al. (2000)) have demonstrated the low-dimensionality of the turbulent wake in flow past a circular cylinder. These results were based on direct numerical simulations (DNS), where all spatial scales down to Kolmogorov scale were accurately resolved. We want to compare the dimensionality of the dynamics in the three ocean regions we simulated with ROMS, HOPS and FVCOM with that of the cylinder flow in order to gain some insight on its complexity. In Fig. 8 we plot the eigenspectra of the cylinder flow at $Re = 10,000; 3,900;$ and 185 together with the ocean eigenspectra corresponding to the *long-term* simulations for the LaTTE, Mass Bay, and Gulf of Maine data.[1] Based on the relative rapid decay of the ocean eigenspectra for all three cases, it is clear that the dynamics that ROMS, HOPS and FVCOM simulate is low-dimensional and certainly much less complex than the cylinder dynamics at $Re = 10,000$ for which there is significant energetic contribution from the high modes unlike the ocean dynamics. To appreciate also the distinction between the *short-term* and *long-term* ocean dynamics, we plot in Fig. 9 the eigenspectra for the short- and long-term simulations of Mass Bay; at the

[1] We note that at $Re = 185$ the cylinder wake is three-dimensional but laminar whereas at $Re = 3,900$ it is turbulent with a small inertial range; at $Re = 10,000$ the wake is fully turbulent although the boundary layer around the cylinder is still laminar.

40th mode the energy content for the former is about two orders of magnitude smaller than the energy content of the latter. We will analyze these two cases in more detail in section 4.

These and other results not presented here suggest that in the *current* generation of mesoscale ocean models the dynamics *captured* is low-dimensional, and hence reduced-order modeling is possible. This, in turn, implies that both the data assimilation problem as well as the forecasting problem can be reformulated in order to exploit the efficiency of low-dimensional representations, specifically using the POD modes. While this finding is not new, the actual low value of the dimensionality, especially for the short-term dynamics compared to the periodic vortex shedding of the turbulent wake, is an intriguing result. It does not reflect the *real* ocean dynamics, which is clearly much more complex than the cylinder dynamics; after all, the Reynolds number of the flow around a buoy (used possibly for measurements) at modest current is larger than $Re = 10,000$! However, it points to the limitations of the physical modeling in the current generation of mesoscale ocean codes and their inability to resolve accurately energetic contributions beyond a handful of modes. One could possibly argue that this is also a limitation of the coarse spatio-temporal resolution but at least for the results we presented here, see Figs. 3 and 7, this does not seem to be the case. In our view, the problem lies in the modeling of the viscous terms, and future reformulations of turbulence modeling in the framework of large-eddy simulations (LES) – using petaflop resources – will improve the accuracy of simulated ocean dynamics greatly. [2] Such LES-based ocean models may require new data assimilation schemes.

3 POD-based Sensor Placement

The simulation results we obtained from three different codes used currently for regional ocean modeling suggest that only a few POD modes are required to describe the essential ocean dynamics. This, in turn, can be exploited in many different ways; here, we show how to use it to advance the adaptive sampling problem. To this end, instead of sampling the physical structures we will sample the POD spatial modes, but only the most dominant ones. We then recast the reconstruction of the ocean state from a limited number of measurements as a *gappy data problem* and we employ the POD formulation to deal with it. This POD extension was first proposed by (Everson and Sirovich (1995)) in image processing applications. (Venturi and Karniadakis (2004)) have improved its robustness by a modification that ensures that the

[2] This observation is supported by the fact that the aforementioned cylinder eigenspectrum at $Re = 10,000$, simulated not with DNS but based on Reynolds-averaged Navier-Stokes (RANS), resembles the eigenspectrum of Fig. 8 at $Re = 185$.

accuracy is independent of the initial guess required in filling the originally incomplete data. POD-based reconstruction for oceanographic data was also developed independently by (Beckers and Rixen (2003)) and (Alvera-Azcárate et al. (2005)). Also, (Qiu and Chou (2006)) proposed a method called four-dimensional data assimilation method based on SVD, somewhat similar to the method by (Everson and Sirovich (1995)). (Gunes et al. (2006)) compared the POD for reconstructing gappy fields to Kriging interpolation (Chiles and Delfinger (1999)) and found it superior for data sets with high temporal resolution.

3.1 POD-based Reconstruction

A three-dimensional field $\mathbf{u}(\mathbf{x},t)$ can be represented by $\mathbf{u}(\mathbf{x},t)=\sum_{k=1}^{\infty}\mathbf{a}^k(t)\mathbf{\Phi}^k(\mathbf{x})$, or by employing a truncated expansion it can be approximated as

$$\mathbf{u}(\mathbf{x},t)\approx\sum_{k=1}^{K}\mathbf{a}^k(t)\mathbf{\Phi}^k(\mathbf{x}), \tag{1}$$

where K is the number of the basis functions $\mathbf{\Phi}^k(\mathbf{x})$, and $\mathbf{a}^k(t)$ are the time-dependent coefficients. Let us define the gappy vector $\tilde{\mathbf{u}}$, which is the point-wise product of a mask vector $\mathbf{m}$ and the complete vector $\mathbf{u}$. Defining an intermediate solution vector $\tilde{\mathbf{u}}_{\mathcal{K}}$ that uses the *existing* spatial POD functions $\mathbf{\Phi}^k(\mathbf{x})$, we obtain the expansion form

$$\tilde{\mathbf{u}}_{\mathcal{K}}=\sum_{k=1}^{K}\mathbf{b}^k(t),\mathbf{\Phi}^k(\mathbf{x}) \tag{2}$$

defined by the unknown coefficients $\mathbf{b}^k(t)$. Minimization of the error defined between the gappy vector $\tilde{\mathbf{u}}$ and the intermediate solution $\tilde{\mathbf{u}}_{\mathcal{K}}$ results in the best possible time coefficients $\mathbf{b}^k$ in the mean energy sense (L2-norm). The error can be defined in the gappy norm as

$$\mathcal{E}=||\tilde{\mathbf{u}}-\tilde{\mathbf{u}}_{\mathcal{K}}||_m^2, \tag{3}$$

where the subscript m denotes that this is not the standard L2-norm. By defining the gappy inner product $(u,v)_m=((m\cdot u),(m\cdot v))$ ($\cdot$ as point-wise multiplication), the corresponding gappy norm $||v||_m^2=(v,v)_m$ can be defined. We note that the POD functions $\mathbf{\Phi}(\mathbf{x})$ are no longer orthogonal with respect to the gappy inner product. We can write Eq. (3) as

$$\mathcal{E}=||\tilde{\mathbf{u}}||_m^2-2\sum_{k=1}^{K}\mathbf{b}^k\Big(\tilde{\mathbf{u}},\mathbf{\Phi}^k(\mathbf{x})\Big)_m+\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbf{b}^i\mathbf{b}^j\Big(\mathbf{\Phi}^i(\mathbf{x}),\mathbf{\Phi}^j(\mathbf{x})\Big)_m, \tag{4}$$

and by differentiating with respect to $\mathbf{b}^n(t)$, the error-minimization problem leads to the equation

$$\frac{\partial \mathcal{E}}{\partial \mathbf{b}^n} = -2(\tilde{\mathbf{u}}, \mathbf{\Phi}^n)_m + 2\sum_{k=1}^{K} \mathbf{b}^k (\mathbf{\Phi}^k, \mathbf{\Phi}^n)_m = 0. \tag{5}$$

This is a linear system of equations in matrix form for the coefficient $\mathbf{b}^k$ in the form

$$\mathbf{M}\mathbf{b} = \mathbf{f} \tag{6}$$

where the matrix entries are defined as $M_{ij} = (\mathbf{\Phi}^i, \mathbf{\Phi}^j)_m$ and the right-hand-side matrix can be computed from $\mathbf{f}_i = (\tilde{\mathbf{u}}, \Phi^i)_m$, so $\mathbf{f}_i$ represents the projection of the measured data onto the available eigenmodes. We note that the matrix $\mathbf{M}$ is non-singular – as it may be the case in the original method (Everson and Sirovich (1995)) – because the eigenmodes are not obtained from gappy snapshots but rather from the output of the ocean model code. Having obtained the time coefficients $\mathbf{b}^k$, we can find the temporary solution through Eq. (2) and then repair the gappy data $\tilde{\mathbf{u}}$ with new information in $\tilde{\mathbf{u}}_{\mathcal{K}}$ for the missing points.

We need, however, to determine the *criteria* for the optimum placement of the sensors available to sample the POD modes. In (Cohen et al. (2003)) this problem was considered for unsteady flow past a circular cylinder and the sensors were placed at the extrema of the POD modes. This, of course, assumes that one knows these modes in advance or in practice one can approximate (e.g., by extrapolation) these modes from another (e.g., near-by) state. An alternative approach in selecting optimum locations was presented in (Willcox (2005)), where the properties of the linear system in the modified POD formulation were involved; a similar approach was also used in (Mokhasi and Rempfer (2004)). To make this point more clear, we consider the linear system of equations (equation (6)), and we solve for the unknown coefficients $\mathbf{b}^k$ with the forcing $\mathbf{f}_i$ obtained from the measured data. We note that the matrix $\mathbf{M}$ is the *identity* matrix in the case of *complete* data due to orthonormality and its condition number $\kappa(M) = 1$ in that case. However, for gappy data $\kappa(M) > 1$, and hence an "optimization" problem can be set up, where the locations of sensors are obtained so that $\kappa(M)$ is minimized, hence approaching closer the *complete* data case (Willcox (2005)). The hypothesis here is that the closer the value of $\kappa(M)$ is to unity the more accurate the reconstructed field will be. Specifically, (Willcox (2005)) proposed the following greedy algorithm to solve this (non-trivial) optimization problem:

- (i) Consider placing the first sensor: loop over all possible placement points, evaluate $\mathbf{M}$ for each point, and choose the point that minimizes $\kappa(\mathbf{M})$.
- (ii) With the first location set, loop over all possible remaining placement points. For each point, update the mask vector, evaluate $\mathbf{M}$, and choose the

point that minimizes $\kappa(\mathbf{M})$.
- (iii) Repeat step (ii) for all remaining sensor locations.

We have modified the above algorithm in two ways: First, we have changed step (i) so that instead of a single sensor we always start with K number of sensors, i.e., equal to the number of retained POD modes in the expansion. This is required as the matrix $\mathbf{M}$ is *rank-deficient* when the number of sensors is smaller than K and hence the condition number will take an infinite value. Specifically, we place K sensors at the *extrema* of the POD modes as initial conditions and solve for all possible combinations, with the objective of minimizing $\kappa(M)$. Second, after step (iii) we pursue a further optimization by re-locating one-at-a-time all the sensors, successively starting from the first one. We found that this additional step may lead to even smaller values of condition number, see results below. We note that we search all the grid points for this optimization problem. Other approaches are also possible, e.g. using a combined steepest descent-simulated annealing employed in (Mokhasi and Rempfer (2004)).

3.2 Unsteady flow past a cylinder

To investigate which criterion we should be using in placing the sensors for sampling the POD modes, we consider a benchmark problem, namely two-dimensional flow past a circular cylinder, and perform extensive tests. In particular, we present a comparative study on the sensor locations using two different criteria: (1) the locations of the extrema of POD modes according to (Cohen et al. (2003)), and (2) the locations that minimize the condition number $\kappa(M)$ of the matrix in equation (6) according to (Willcox (2005)). To this end, we consider two-dimensional unsteady flow past a circular cylinder. We performed high-resolution spectral/*hp* element [3] simulations at $Re = 100$ for which a time-periodic state is achieved. The POD modes are extracted from snapshots taken over one time period; details can be found in (Ma et al. (2000)).

We assume that we have 12 sensors for the limited number of measurements, and the question is which POD modes to sample and how. For simplicity, we also assume that we will employ only *four* POD modes for the reconstruction. We have investigated more than 20 cases to determine the best possible sensor configurations but here we present only the 10 most representative ones. In general, for each configuration we start with the extrema of the POD modes (using different variables, e.g., the velocity components u, v or the total velocity $U = \sqrt{u^2 + v^2}$ or the temperature T). We then search for locations

[3] The spectral/*hp* element method is a high-order finite element method that employs spectral polynomial as trial basis, see Karniadakis and Sherwin (2005).

that minimize $\kappa(M)$ and evaluate the error of the reconstructed field using as reference solution the results from the full simulation. We summarize the 10 cases in Table 3. In case 1 we do not use the greedy algorithm but instead we place the sensors at the extrema of the POD modes of the total velocity U with the following sensor distribution (2,2,4,4) for the modes (1,2,3,4), respectively. In other words, here we place more sensors in the higher modes (3 & 4). However, we have investigated other configurations, e.g. (4,4,2,2) or (3,3,3,3). Specifically, we examined another nine cases for which the sensors are placed at the extrema of different POD modes and with different number of sensors per mode; the results are somewhat similar to case 1 so we do not present them here. In case 2 all the sensors are placed along the centerline. For the other cases (3-10) the aforementioned modified *greedy algorithm* was used to find the best possible locations. For cases 3 and above we have used different initial configurations as follows. Case 3 uses case 2 as initial configuration and then a greedy algorithm is run (i.e., the last part only of the relocation part of the algorithm). Similarly, case 4 uses case 3 as initial configuration, and so on. We identified 100 extrema in the POD modes and starting with these as possible initial locations we had to perform extensive searches. Subsequently, we re-located all sensors, one-at-a-time in an iterative fashion, to look for possibly better sensor locations.

In order to compare the various cases we define the *reconstruction error* as

$$e_j = \sqrt{\frac{1}{N}\sum_{1}^{N}\frac{\int_\Omega(\tilde{\mathbf{u}}_i - \mathbf{u}_i)^2}{\int_\Omega \mathbf{u}_i^2}},$$

where N is the number of snapshots, and $\tilde{\mathbf{u}}$ and $\mathbf{u}$ denote reconstructed and reference fields, respectively. We should also distinguish between the *measurement error* and the *total error* (measurement plus truncation). The latter reflects the inaccuracies committed due to the fact that we only employed *four* POD modes instead of the total number of POD modes available. Hence, we use the subscript j that takes the value of 1 or 2 for the *total* or *measurement* errors, respectively.

In Table 4 we summarize the results for the aforementioned 10 cases. Overall, we have observed that a small condition number $\kappa(M)$ corresponds to small reconstruction errors e_2 (and hence e_1) but the absolutely smallest $\kappa(M)$ does not correspond to absolutely smallest e_2. We note that in case 2, with all 12 sensors along the centerline, we obtain a very large condition number and correspondingly a very large reconstruction error. Cases 4, 7 and 10 have different $\kappa(M)$ but about the same error e_2. In case 1 (sensors on the extrema of POD modes) – for which no optimization solution is required for $\kappa(M)$ - we have the second smallest error e_2 although its condition number is well above 1. We have observed similar trends by considering the POD modes of other variables (e.g., temperature or vorticity) and placing the sensors as above but following

the contour levels of these state variables. We have concluded that configurations corresponding to low condition number should be preferred for low reconstruction error. However, placing the sensors at the extrema of the POD modes of the most important state variable leads to both low reconstruction errors and low condition number without the extra overhead associated with the search/optimization algorithm.

In Fig. 10 we plot the first four POD modes along with the location of the sensors for case 1; the sensors are placed at the extrema of the POD modes. In Fig. 11 we plot the locations of 12 sensors for case 3, the coordinates of which were computed based on the greedy algorithm. Comparing the results shown in the two figures, we see that the sensors are placed in different positions although the reconstruction error is of about the same magnitude, i.e., 6-7%. This, of course, indicates that there is no unique solution to this problem and hence the most efficient method should be used in practice, i.e., the extrema of the POD modes.

In summary, the general conclusion we draw from these results is that the extrema of POD modes are very good locations, if not optimum, to place the sensors. At these locations, the condition number $\kappa(M)$, becomes small but not minimum. In fact, minimization of $\kappa(M)$ does not necessarily lead to the smallest reconstruction field.

Next, we present results for the Mass Bay using the most efficient method, i.e., selecting the extrema of POD modes as sensor locations.

4 Massachusetts Bay: Results and Discussion

We now apply the modified POD framework in sampling Mass Bay. We will investigate the level of reconstruction errors for a different number of retained POD modes (e.g., different truncations) and also for different number of sensors for both *short-term* (8-days) and *long-term* (47-days) simulations. Following the findings from our cylinder investigation, we will place the sensors at the *extrema* of the POD modes; a typical result is shown in Fig. 12. We still have many choices regarding the sensor distribution per mode; here, we will only show a subset of our results. We use as guiding modes the ones corresponding to the total velocity $U(=\sqrt{u^2+v^2})$ and also the temperature. The sensor configurations for all cases are given in Table 5. Each box in the table contains information on how the sensors are distributed per mode; for example, $2x-2x-4x-4x$ for the four-modes in case 1 implies that (2,2,4,4) sensors are used for modes (1,2,3,4), respectively, if $x=1$ or (4,4,8,8) if $x=2$, depending on the number of the available sensors (12 and 24, respectively).

We will measure the time-averaged reconstruction errors using the following definition:

$$e_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\int_\Omega(\tilde{\mathbf{u}}_i - \mathbf{u}_i)^2}{\int_\Omega(\mathbf{u}+\mathbf{u}_{avg})_i^2}}, \quad \text{for} \quad j = 1, 2,$$

where the subscript "i" denotes snapshot number. The reference state $\mathbf{u}$ is meant to be a complete HOPS solution for $j = 1$ and reconstructed POD solution for $j = 2$ with as many modes (4, 6, 8, 16) as we indicate in each case in Table 5. We note that in the definition of the errors e_1, e_2 we normalize by the total magnitude, i.e., we include the velocity $\hat{\mathbf{u}}_{avg}$.

4.1 Total velocity

We first present errors for the *short-term* simulation using 4, 6 and 8 POD modes in the expansion. For the 4-modes case, the results are given in Fig. 13. The error e_1 is about 10.5% for 12 sensors and 9.5% for 24 sensors. The corresponding truncation error $\tilde{e}_1$ is about 8.5% and hence it dominates. In Fig. 14, we show results for the 6-modes case and we see that the e_1 error is reduced to about 7.2% for 24 sensors and 6.8% for 48 sensors. Clearly, using twice as many sensors is *not* an effective way to reduce the error since the truncation error dominates. In fact, we can use only 40 sensors and instead increase the number of modes to 8 leading to a greater error reduction, i.e., the error e_1 decreases to about 4.5%, see Fig. 15. More specifically, the error e_2 that excludes the truncation error is only 2% in this case; however, doubling the number of sensors from 40 to 80 decreases the error slightly to about 1.5% (for cases 1 and 2 only).

What we have seen so far from the short-term simulation results is that we have to look for the appropriate combination of number of modes and number of sensors for effective error reduction and efficient utilization of the resources. We will apply the same approach to results from long-term simulations, where we will now use 4, 8, and 16 modes to reconstruct new fields and we also assume that we can use (12,24) (40, 80) and (144,288) sensors, respectively.

In Fig. 16, reconstruction errors for the 4-modes *long-term integration* are shown. The corresponding truncation error is $\tilde{e}_1 = 31.7\%$, which is a lower limit for the reconstruction errors e_1. The error e_1 for all three cases is about 36% for 12 sensors; increasing the total sensors to 24, e_1 is lowered by only 2%. These are large errors although the corresponding values of the condition number are low, e.g., 7.10 and 8.12 for case 3 corresponding to 12 and 24 sensors. For the 8-modes case (results not shown here), the difference between e_1 and the truncation error $\tilde{e}_1$ has dropped to about 3% with 40 sensors and about 2% with 80 sensors. In comparison with the 4-modes results, the error dropped about 16% more for the 8-modes case. Doubling the number of modes

to 16, see Fig. 17, yields an e_1 error of about 21% which is only 2% above the truncation error.

In summary, by comparing the short-term and long-term simulation we can appreciate how the reconstruction errors are depended on the dimensionality of the system. For adaptive sampling in mesoscale regions, the short-term results are more relevant, but we include the long-term results as an extreme case. However, even in this case, the reconstructed field has all the salient features of the original field as it is shown in the comparison in figure 18.

4.2 Temperature

We have also applied the POD method in reconstructing the temperature field for short-term and long-term simulations. The reconstructed fields for temperature are much more accurate than the velocity due to the lower dimensionality of the temperature field. Specifically, in the *short-term* simulation, only four modes contain 97.9% of total energy (truncation error about 0.7%) whereas in the *long-term simulation* four modes contain 65.7% of the total energy (truncation error 5.0%); this is a larger percentage than the corresponding 53.2% of total velocity energy content. Using 12 sensors, the reconstruction error e_1 is about 1% and 6% for the short-term and long-term simulations, respectively. On the other hand, if we employ 16 POD modes, the truncation error e_1 for the long-term simulation drops to 2.4% and the reconstruction error (using 144 sensors) drops to about 2.6%; it is negligible for the short-term simulation.

Here we present detailed results for eight modes as shown in Fig. 19 for the short-term and long-term simulations. In the former, the truncation error is 0.3% while in the latter is 3.6%. We use 40 sensors for each simulation distributed according to the aforementioned three different cases. We see that for the short-term simulation, the reconstruction error in the temperature field is slightly above the truncation error and no greater than 0.5% for all three cases whereas for the long-term simulation the reconstruction error is about 4% due to the larger truncation error. The condition number $\kappa(M)$ is smallest for case 3 for both short-term and long-term simulations (5.37 and 10.59, respectively).

While the previous plots provide a quantitative assessment of the reconstruction errors, in Fig. 20 we present a more global but qualitative comparison of the temperature field between the original data ("the ground truth") and the reconstructed field at one time instant. The agreement is very good and it is typical for all other time instants as well.

5 Summary and Discussion

In this paper we have addressed the question of how to identify efficiently the "best" locations for sampling the ocean state for the most accurate reconstruction of the velocity, temperature and salinity fields from a limited number of measurements, in the context of fast adaptive sampling. To this end, we first performed simulations of three different regions using the mesoscale ocean models ROMS, HOPS and FVCOM and analyzed their outputs using proper orthogonal decomposition (POD). Specifically, we investigated both short-term and long-term simulation outputs and obtained the corresponding POD eigenvalues and eigenmodes. Our results indicate that the "numerical" ocean is low-dimensional and hence a handful of modes is sufficient to describe the essential dynamics. In particular, we demonstrated that even the long-term dynamics simulated with HOPS for Mass Bay has dimensionality lower than the dynamics of the turbulent wake at $Re = 10,000$ exhibiting periodic shedding. This, in turn, suggests that in reconstructing an ocean state it is advantageous to capture the most energetic modes, i.e., perform sampling in *modal* space rather than in physical space. We employed POD and investigated two different strategies for sensor placement. In the first approach, the extrema of the POD modes are selected as the sensor locations whereas in the second one, an iterative procedure is set up that aims to minimize the condition number of a matrix involved in the POD approach. We found that the latter approach is more expensive than the former and does not lead to any significant accuracy gains. These are not the only possible optimization approaches, and one can use the reconstruction error directly as alternative cost function or set up different minimization procedures using other sets of sampling points, e.g. see (Nguyen et al. (2008)). As stated in (Nguyen et al. (2008)), there can be several sets of "best" points as uniqueness is not guaranteed. For adaptive sampling in forecasting the ocean state, in particular, the simplest and most efficient approach is the desirable one.

Based on the extrema of POD modes, we reconstructed the velocity field for the Mass Bay for 8-days and 47-days simulations. We found that for the 8-days simulation even a relatively small number of sensors (e.g., 12) can give an accurate reconstructed velocity field. However, for a 47-days simulation even for a large number of sensors the reconstruction errors are large. The latter case is not representative of adaptive sampling but we included it in order to investigate how the POD approach behaves in that extreme limit. It also points to the connection we attempted to make between the dimensionality of the system and the particular scheme of data assimilation that needs to be employed.

The accuracy of our method is shown to strongly depend on the number of employed POD modes, and hence it is important how to distribute the

available sensors among the POD modes to be sampled. Another question is which state variables to use in a *multi-physics* simulation involving velocity, temperature, salinity and possibly chemistry or biology. We have found that the *total velocity* $U = \sqrt{u^2 + v^2}$ (employed in this paper) but also the *vorticity* are the best variables. In particular, we simulated scenarios where the extrema of the temperature POD modes were selected as the measurement locations, and we found that the errors in all other fields increased compared to the cases where the extrema of the total velocity or vorticity POD modes were employed. This is expected given that the POD eigenspectra of the velocity field show a broader distribution of the energy (i.e., higher dimensionality) compared to the temperature or the salinity eigenspectra, and hence higher dimensionality. From the practical standpoint, given the currently available resources for multi-field measurements, e.g. CTD, it is therefore better to take measurements at the extrema of the total velocity POD modes even though these may not be the theoretically guaranteed best locations for the other fields.

Another open question is *which POD modes* to use in the reconstruction. Here in order to focus on the reconstruction error only, we employed POD modes from the full simulation. In practice, such modes will not be available and hence some approximations are required. In the adaptive sampling context, as we advance the ocean computer model in time, such POD modes can be constructed from the snapshots obtained at previous time steps, hence there will be a time lag due to such extrapolation. In preliminary tests we performed we found differences less than 5%, in agreement also with the results of (Mokhasi and Rempfer (2004)). In general, given the small time steps involved in the mesoscale ocean models we do not anticipate any major error contributions from this time extrapolation. Specifically, let us assume that we have a time window of two to three hours to deploy our sensors; using snapshots from the previous hour to obtain the POD modes will not significantly increase the reconstruction errors. This is also justified by our results for short-term integration that show extremely small errors in the short-term dynamics runs.

Another issue that requires a physical explanation is why the extrema of POD modes are such important locations, and specifically how are these locations related to the uncertainty fields employed, e.g., in HOPS via the ESSE system (Lermusiaux (2001)). In work not presented here, we performed direct comparisons with the ESSE approach for the Mass Bay, and we found the the POD extrema coincide or are very close to the locations of maximum uncertainty employed in the ESSE data assimilation scheme. From the approximation theory standpoint, one can draw an analogy with non-Fourier spectral theory where the maxima of the Chebyshev polynomials are the collocation points (the so-called Gauss-Lobatto points), based on which the most accurate approximations are guaranteed – we refer here to the so called "minimax theorem", see (Gottlieb and Orszag (1977)). Clearly the POD modes

are not necessarily polynomials but they can be approximated by spectral polynomials.

Finally, we emphasize that the "measured data" in this study are of "infinite" accuracy, however, in practice we have to incorporate the uncertainty in our measurements and propagate it through the computer ocean model along with other parametric uncertainties of the model. Clearly, further systematic studies are required to resolve these issues before the use of our proposed approach in future adaptive sampling experiments.

Acknowledgement

We would like to thank the three anonymous referees and the Associate Editor Dr. S. Griffies whose insightful comments helped us improve the quality of our manuscript. This work was supported by ONR (Dr. T.F. Swean) and by the MIT Sea Grant program. We would like to thank Prof. P. Leirmusiaux (MIT), Prof. D. Haidvogel & Dr. J. Levin (Rutgers University) and Prof. C. Chen (University of Massachusetts at Dartmouth) for providing the simulation data and for many helpful discussions.

References

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J. M., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic sea surface temperature. Ocean Modelling 9, 325–346.

Arango, H., Moore, A., Miller, A., Cornuelle, B., Lorenzo, E., Neilson, D., 2003. The ROMS tangent linear and adjoint models: A comprehensive ocean prediction and analysis system. Report, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ.

Aubry, N., Guyonnet, R., Stone, E., 1991. Spatio-temporal analysis of complex signals: Theory and applications. J. Stat. Phys. 64, 683–739.

Beckers, J., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic datasets. J. Atmos. Ocean Tech. 20 (12), 1839–1856.

Bekooz, G., Holmes, P., Lumley, J., 1993. The proper orthogonal decomposition in the analysis of turbulent flows. Ann. Rev. Fluid Mech. 25, 539–575.

Buehner, M., Malanotte-Rizzoli, P., 2003. Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation. Journal of Geophysical Research 108, 3192.

Chiles, J.-P., Delfinger, P., 1999. Geostatistics, Modeling Spatial Uncertainty. Wiley Series in Probability and Statistics.

Choi, B., Wilkin, J. L., 2007. The effect of wind on the dispersal of the Hudson river plume. J. Phys. Ocean. 37, 1878–1897.

Cohen, K., Siegel, S., McLaughlin, T., 2003. Sensor placement based on proper orthogonal decomposition modeling of a cylinder wake. AIAA Paper 2003-4259.

D'Andrea, F., Vautard, R., 2001. Extratropical low-frequency variability as a low dimensional problem. Part I: A simplified model. Quart. J. Roy. Meteor. Soc. 127, 1357–1375.

Dickey, T., 2003. Emerging ocean observations for interdisciplinary data assimilation systems. Journal of Marine Systems 40-41, 5–48.

Dong, S., Karniadakis, G., Ekmekci, A., Rockwell, D., 2006. A combined DNS-PIV study of the turbulent near-wake. J. Fluid Mech. 569, 185–207.

Everson, R., Cornillon, P., Sirovich, L., Webber, A., 1995. An empirical eigenfunction analysis of sea surface temperatures in the North Atlantic. J. Phys. Ocean. 27(3), 468–479.

Everson, R., Sirovich, L., 1995. The Karhunen-Loève procedure for gappy data. J. Opt. Soc. Am., A 12(8), 1657–1664.

FVCOM, URL. http://fvcom.smast.umassd.edu.

Gottlieb, D., Orszag, S., 1977. Numerical analysis of spectral methods: Theory and Applications. SIAM-CMBS, Philadelphia.

Gunes, H., Sirisup, S., Karniadakis, G., 2006. Gappy data: To Krig or not to Krig? J. Comp. Phys. 212, 358–382.

Hendricks, J., Leben, R., Born, G., Koblinsky, C., 1996. Empirical orthogonal function analysis of global TOPEX/POSEIDON altimeter data and implications for detection of global sea rise. J. Geophys. Res. 101, 14131–14145.

HOPPS-ESSE, 2003. HOPPS-ESSE contributions to the Autonomous Sampling Network II (AOSN-II) field exercise. http://www.deas.harvard.edu/leslie/aosnii/index.html.

HOPS, URL. http://oceans.deas.harvard.edu/HOPS/HOPS.html.

Houseago-Stokes, R., 2000. Using optimal interpolation and EOF analysis on North Atlantic satellite data. International WOCE Newsletter 28, 26–28.

Karniadakis, G., Sherwin, S., 2005. Spectral/hp element methods for CFD. Oxford University Press, 2nd edition, New-York, NY, USA.

LATTE, URL. http://marine.rutgers.edu/cool/latte/.

Lermusiaux, P., 2001. Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay. J. Marine Systems 29/1-4, 385–422.

Lermusiaux, P., Malanotte-Rizzoli, P., Stammer, D., Carton, J., Cummings, J., Moore, A., 2006. Progress and Prospects of U.S. Data Assimilation in Ocean Research. Oceanography 19, 172–183.

Ma, X., Karamanos, G., Karniadakis, G., 2000. Dynamics and low-dimensionality of the turbulent near-wake. J. Fluid Mech. 410, 29–65.

Ma, X., Karniadakis, G., 2002. A low-dimensional model for simulating 3d cylinder flow. J. Fluid Mech. 458, 181–190.

Mokhasi, P., Rempfer, D., 2004. Optimized sensor placement for urban flow

measurement. Physics of Fluids 16(5), 1758.
Nguyen, N., Patera, A., Peraire, J., 2008. A "best points" interpolation method for efficient basis approximation of parametrized functions. International Journal of Numerical Methods in luids 73(4), 521.
Pedder, M., Gomis, D., 1998. Application of EOF Analysis to the spatial estimation of circulation features in the ocean sampled by high-resolution CTD samplings. J. Atmos. Ocean Tech. 15(4), 959–978.
Preisendorfer, W., Mobley, C. D., 1988. Principal component analysis in meteorology and oceanography. Elsevier.
Qiu, C., Chou, J., 2006. Four-dimensional data assimilation method based on SVD: Theoetical aspect. Theor. Appl. Climatol. 83, 51–57.
Rempfer, D., 2003. Low-dimensional modeling and numerical simulation of transition in simple shear flow. Ann. Rev. Fluid Mech. 35, 229–265.
Robinson, A., Glenn, S., 1999. Adaptive sampling for ocean forecasting. Naval Research Reviews 51(2), 27–38.
ROMS, URL. http://www.myroms.org.
Rutgers Group, URL. http://marine.rutgers.edu.
Sirovich, L., 1987. Turbulence and the dynamics of coherent structures, Parts I, II and III. Quart. Appl. Math. XLV, 561–590.
Venturi, D., Karniadakis, G., 2004. Gappy data and reconstruction procedures for flow past a cylinder. J. Fluid Mech. 519, 315–336.
Wilkin, J., Zavala-Garay, J., Arango, H., 2008. Predictability of mesoscale variability in the East Australia current system given strong constraint data assimilation. AGU/ASLO Ocean Sciences Meeting, Orlando, Florida.
Wilkin, J., Zhang, W., 2006. Modes of mesoscale sea surface height and temperature variability in the East Australian Current. J. Geophys. Res. 112, C01013.
Willcox, K., 2005. Unsteady flow sensing and estimation via the gappy proper ortogonal decomposition. Computer & Fluids 35(2), 208–226.
Zang, X., Malanotte-Rizzoli, P., 2003. A comparison of assimilation results from the Ensemble Kalman filter and the Reduced-Rank Extended Kalman filter. Nonlinear Processes in Geophysics 10, 6477–6491.

Figure Captions

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

Figure 2: New Jersey coast: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for 2.5-days (left) and 25-days (right) simulations.

Figure 3: New Jersey coast: Time resolution effects on the energy spectra of horizontal velocity (u, v) vector for 2.5-days (left) and 25-days (right) simulations.

Figure 4: Mass Bay: Contours of the first POD mode (dimensionless temperature) and domain.

Figure 5: Mass Bay: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for 8-days (left) and 47-days simulations.

Figure 6: Gulf of Maine: Contours of the first POD mode (temperature) and domain.

Figure 7: Gulf of Maine. Left: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for Gulf Of Maine. Right: Energy spectra of horizontal velocity vector at two different grid resolutions.

Figure 8: Comparison of eigenspectra between the turbulent wake and the long-term ocean dynamics of three regions.

Figure 9: Comparison of eigenspectra between the turbulent wake and the short-term and long-term ocean dynamics of Mass Bay.

Figure 10: Contours of first four POD modes of the total velocity U and sensor locations for case 1. Mode 1: upper-left; Mode 2: upper-right; Mode 3: lower-left; Mode 4: lower-right.

Figure 11: Contours of streamwise velocity u and sensor locations for case 3.

Figure 12: Mass Bay: Schematic of best sensor locations for Case 1 (see Table 5: 4 modes, 12 sensors). Only a slice at the surface of the first POD mode is shown; different symbols correspond to different modes. The extrema are located at some depth from the surface. The contours represent the second POD mode of the temperature.

Figure 13: Short-term simulation, 4-modes: Time averaged errors for total velocity. The total of *four* modes contain 92.6% of total energy and the corresponding truncation error is $\tilde{e}_1 = 8.5\%$. For 12 sensors $\kappa(M) = 14.82, 10.54, 9.55$ for Cases 1,2 and 3, respectively.

Figure 14: Short-term simulation, 6-modes: Time averaged errors for total velocity. The total of *six* modes contain 96.1% of total energy and the corresponding truncation error is $\tilde{e}_1 = 6.2\%$. For 24 sensors $\kappa(M) = 18.92, 17.74, 16.54$ for Cases 1,2 and 3, respectively.

Figure 15: Short-term simulation, 8-modes. Time averaged errors for total velocity. The total of *eight* modes contain 98.1% of total energy and the truncation error is $\tilde{e}_1 = 4.3\%$. For 40 sensors $\kappa(M) = 19.15, 20.37, 19.94$ for Cases 1,2 and 3, respectively.

Figure 16: Long-term simulation, 4-modes. Time averaged errors for total velocity. The total of *four* modes contain 53.2% of total energy and the truncation

error is $\tilde{e}_1 = 31.7\%$. For 12 sensors $\kappa(M) = 20.73, 11.33, 7.10$ for Cases 1,2 and 3, respectively.

Figure 17: Long-term simulation, 16-modes. Time averaged errors are for total velocity. The total of *sixteen* modes contain 83.1% of total energy and the truncation error is $\tilde{e}_1 = 19.3\%$. For 144 sensors $\kappa(M) = 13.42, 13.92, 11.78$ for Cases 1,2 and 3, respectively.

Figure 18: Comparison between the original (left) and the reconstructed (right) total velocity field for the 24th snapshot in the long-term simulation.

Figure 19: Time averaged errors for temperature; 8-modes. Both short-term and long-term results are shown using 40 sensors.

Figure 20: Comparison between the original (left) and the reconstructed (right) temperature field for the 36th snapshot in the short-term simulation.

Table 1. Mean (First) and subsequent eight eigenmodes for New Jersey Coast ($^1 = 2.5$ days, $^2 = 25$ days), Massachusetts Bay ($^1 = 8$ days, $^2 = 47$ days) and Gulf of Maine simulations.

Mode No:	NJ Coast ^{1}			NJ Coast ^{2}			Mass. Bay ^{1}			Mass. Bay ^{2}			Gulf of Maine		
	Velocity	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*
First Mode	30.52 %	99.94 %	99.99 %	31.82 %	99.72 %	99.90 %	62.90 %	99.32 %	99.99 %	51.62 %	99.13 %	99.99 %	68.95 %	99.82 %	99.98 %
Second Mode	23.34 %	0.02 %	0.01 %	20.26 %	0.14 %	0.07 %	14.57 %	0.43 %	< 0.01 %	15.11 %	0.26 %	< 0.01 %	20.77 %	0.04 %	0.01 %
Third Mode	12.38 %	0.01 %	< 0.01 %	12.96 %	0.04 %	0.01 %	8.75 %	0.12 %	< 0.01 %	6.32 %	0.13 %	< 0.01 %	5.95 %	0.03 %	< 0.01 %
Fourth Mode	10.46 %	< 0.01 %	< 0.01 %	10.88 %	0.02 %	0.01 %	6.85 %	0.06 %	< 0.01 %	5.08 %	0.09 %	< 0.01 %	0.68 %	0.03 %	< 0.01 %
Fifth Mode	8.58 %	< 0.01 %	< 0.01 %	5.68 %	0.01 %	< 0.01 %	2.40 %	0.02 %	< 0.01 %	4.28 %	0.08 %	< 0.01 %	0.43 %	0.01 %	< 0.01 %
Sixth Mode	7.54 %	< 0.01 %	< 0.01 %	4.39 %	0.01 %	< 0.01 %	1.71 %	< 0.01 %	< 0.01 %	2.45 %	0.06 %	< 0.01 %	0.32 %	0.01 %	< 0.01 %
Seventh Mode	1.89 %	< 0.01 %	< 0.01 %	1.91 %	< 0.01 %	< 0.01 %	0.72 %	< 0.01 %	< 0.01 %	1.79 %	0.04 %	< 0.01 %	0.25 %	0.01 %	< 0.01 %
Eighth Mode	1.54 %	< 0.01 %	< 0.01 %	1.72 %	< 0.01 %	< 0.01 %	0.42 %	< 0.01 %	< 0.01 %	1.53 %	0.03 %	< 0.01 %	0.19 %	< 0.01 %	< 0.01 %
Nineth Mode	0.99 %	< 0.01 %	< 0.01 %	1.49 %	< 0.01 %	< 0.01 %	0.37 %	< 0.01 %	< 0.01 %	1.21 %	0.02 %	< 0.01 %	0.16 %	< 0.01 %	< 0.01 %

Table 2. First eight eigenmodes for New Jersey Coast ($^1 = 2.5$ days, $^2 = 25$ days), Massachusetts Bay ($^1 = 8$ days, $^2 = 47$ days) and Gulf of Maine simulations.

Mode No:	NJ Coast ^{1}			NJ Coast ^{2}			Mass. Bay ^{1}			Mass. Bay ^{2}			Gulf of Maine		
	Velocity	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*	*Velocity*	*Temperature*	*Salinity*
First Mode	33.6 %	49.1 %	59.1 %	29.7 %	49.5 %	74.7 %	39.3 %	64.0 %	59.2 %	31.2 %	29.9 %	34.3 %	66.9 %	24.3 %	71.4 %
Second Mode	17.8 %	13.6 %	17.6 %	19.0 %	16.3 %	8.8 %	23.6 %	18.0 %	18.8 %	13.1 %	15.0 %	16.4 %	19.2 %	17.3 %	9.7 %
Third Mode	15.1 %	8.1 %	10.5 %	16.0 %	9.8 %	6.3 %	18.5 %	8.3 %	10.0 %	10.5 %	10.7 %	10.5 %	2.2 %	14.5 %	4.5 %
Fourth Mode	12.3 %	6.0 %	3.7 %	8.3 %	7.4 %	2.6 %	6.5 %	3.4 %	4.4 %	8.9 %	8.8 %	6.3 %	1.4 %	7.6 %	2.9 %
Fifth Mode	10.9 %	5.1 %	1.6 %	6.4 %	3.1 %	1.4 %	4.6 %	2.2 %	2.6 %	5.1 %	6.4 %	5.3 %	1.0 %	6.0 %	2.8 %
Sixth Mode	2.7 %	3.8 %	1.0 %	2.8 %	2.4 %	0.9 %	1.9 %	1.7 %	1.8 %	3.7 %	4.3 %	4.4 %	0.8 %	3.7 %	1.2 %
Seventh Mode	2.2 %	3.3 %	0.9 %	2.5 %	2.0 %	0.7 %	1.1 %	0.8 %	0.9 %	3.2 %	3.1 %	3.7 %	0.6 %	2.8 %	1.2 %
Eighth Mode	1.4 %	2.6 %	0.8 %	2.2 %	1.4 %	0.6 %	1.0 %	0.5 %	0.6 %	2.5 %	2.6 %	2.6 %	0.5 %	1.9 %	0.9 %

Table 3
Sensor location for 12 sensors. In case 1 $\mathbf{x}_i^j$ denotes the coordinate of mode j with i^{th} extrema, with 1 corresponding to largest extremum; the contours of the POD modes of total velocity $U = \sqrt{u^2 + v^2}$ are employed for identifying the extrema.

Sensor No: / Case No:	1	2	3	4	5	6
1	$x = x_1^1$ $y = y_1^1$	$x = x_2^1$ $y = y_2^1$	$x = x_1^2$ $y = y_1^2$	$x = x_2^2$ $y = y_2^2$	$x = x_1^3$ $y = y_1^3$	$x = x_2^3$ $y = y_2^3$
2	$x = 1.0$ $y = 0.0$	$x = 2.0$ $y = 0.0$	$x = 3.0$ $y = 0.0$	$x = 4.0$ $y = 0.0$	$x = 5.0$ $y = 0.0$	$x = 6.0$ $y = 0.0$
3	$x = 0.62$ $y = -0.11$	$x = 18.9$ $y = 0.05$	$x = 4.19$ $y = -1.28$	$x = 1.49$ $y = -0.6$	$x = 1.96$ $y = -1.0$	$x = 2.78$ $y = 1.42$
4	$x = 2.0$ $y = 0.00$	$x = 4.0$ $y = 0.0$	$x = 6.0$ $y = 0.0$	$x = 14.7$ $y = 5.47$	$x = -0.87$ $y = -1.86$	$x = 0.37$ $y = -0.53$
5	$x = 1.25$ $y = -0.2$	$x = 1.38$ $y = 0.11$	$x = 8.30$ $y = 0.57$	$x = 6.52$ $y = -2.29$	$x = 7.48$ $y = -0.08$	$x = 7.05$ $y = -1.21$
6	$x = 2.06$ $y = -2.0$	$x = 2.01$ $y = 0.0$	$x = 2.06$ $y = 2.01$	$x = 0.38$ $y = -0.56$	$x = 3.25$ $y = 0.0$	$x = 13.8$ $y = -1.93$
7	$x = 0.87$ $y = -0.48$	$x = 2.79$ $y = 0.0$	$x = 2.53$ $y = -1.79$	$x = 1.18$ $y = -1.11$	$x = 3.25$ $y = 0.15$	$x = 3.49$ $y = 1.59$
8	$x = 3.12$ $y = 0.74$	$x = 3.15$ $y = -0.75$	$x = 2.31$ $y = 0.62$	$x = 1.41$ $y = -0.24$	$x = 0.96$ $y = 0.01$	$x = 3.36$ $y = -1.29$
9	$x = 3.24$ $y = -0.5$	$x = 3.12$ $y = -0.8$	$x = 2.31$ $y = 0.62$	$x = 8.93$ $y = 0.21$	$x = 12.3$ $y = 0.0$	$x = 3.17$ $y = -1.26$
10	$x = 6.38$ $y = 1.17$	$x = 5.21$ $y = 1.04$	$x = 10.8$ $y = 0.0$	$x = 1.14$ $y = 0.0$	$x = 0.73$ $y = -0.38$	$x = 24.8$ $y = -1.38$

Sensor No: / Case No:	7	8	9	10	11	12
1	$x = x_3^3$ $y = y_3^3$	$x = x_4^3$ $y = y_4^3$	$x = x_1^4$ $y = y_1^4$	$x = x_2^4$ $y = y_2^4$	$x = x_3^4$ $y = y_3^4$	$x = x_4^4$ $y = y_4^4$
2	$x = 7.0$ $y = 0.0$	$x = 8.0$ $y = 0.0$	$x = 9.0$ $y = 0.0$	$x = 10.0$ $y = 0.0$	$x = 11.0$ $y = 0.0$	$x = 12.0$ $y = 0.0$
3	$x = 6.71$ $y = -1.36$	$x = 7.89$ $y = -0.76$	$x = 6.25$ $y = -0.15$	$x = 7.21$ $y = 0.36$	$x = 1.03$ $y = -0.34$	$x = 9.36$ $y = -0.3$
4	$x = 5.46$ $y = -1.19$	$x = 2.42$ $y = -0.46$	$x = 5.47$ $y = 1.24$	$x = 3.27$ $y = -0.26$	$x = 1.56$ $y = 0.63$	$x = 14.2$ $y = 1.47$
5	$x = 1.15$ $y = 0.41$	$x = 2.52$ $y = -0.59$	$x = 1.56$ $y = -0.52$	$x = 3.30$ $y = -0.20$	$x = 1.54$ $y = 0.06$	$x = 8.85$ $y = -1.04$
6	$x = 17.8$ $y = 1.72$	$x = 3.91$ $y = -0.77$	$x = 3.0$ $y = -0.68$	$x = 4.0$ $y = 0.74$	$x = 3.21$ $y = -0.41$	$x = 18.9$ $y = 1.25$
7	$x = 22.4$ $y = 0.70$	$x = 3.92$ $y = -0.77$	$x = 3.0$ $y = -0.68$	$x = 4.0$ $y = 0.74$	$x = 3.21$ $y = -0.41$	$x = 11.0$ $y = 1.89$
8	$x = 2.42$ $y = -0.24$	$x = 22.3$ $y = -0.31$	$x = 9.66$ $y = -0.04$	$x = 6.32$ $y = -0.32$	$x = 9.67$ $y = -0.03$	$x = 7.64$ $y = 0.22$
9	$x = 17.7$ $y = 0.52$	$x = 14.4$ $y = 0.49$	$x = 8.34$ $y = 0.0$	$x = 6.32$ $y = -0.32$	$x = 9.64$ $y = 0.0$	$x = 7.64$ $y = 0.22$
10	$x = 24.1$ $y = -0.78$	$x = 1.26$ $y = -1.44$	$x = 24.2$ $y = 0.54$	$x = 0.51$ $y = 0.1$	$x = 0.63$ $y = -0.45$	$x = 0.51$ $y = 0.0$

Table 4
Errors e_1 (w.r.t. full simulation) and e_2 (w.r.t. POD reconstruction) of the total velocity $U = \sqrt{(}u^2 + v^2)$.

Case No:	Cond. Number $\kappa(M)$:	e_1	e_2
1	4.1932	15.79 %	7.42 %
2	2.3e+08	66283 %	66939.4 %
3	1.0359	15.33 %	6.37 %
4	3.0940	18.66 %	12.49 %
5	1.0276	17.32 %	10.33 %
6	1.5024	19.17 %	13.25 %
7	1.0152	18.38 %	12.05 %
8	1.4815	15.95 %	7.76 %
9	1.0327	16.10 %	8.07 %
10	1.0249	18.16 %	11.71 %

Table 5
Sensor configurations for Mass Bay. In the table-boxes, N x means N times x, with x a multiplier adjusted for the total number of sensors.

Total Modes: / Case No:	Four Modes	Six Modes	Eight Modes	Sixteen Modes
Case 1	2x-2x-4x-4x	2x-2x-4x-4x-6x-6x	2x-2x-4x-4x-6x-6x-8x-8x	2x-2x-4x-4x-6x-6x-8x-8x-10x-10x-12x-12x-14x-14x-16x-16x
Case 2	3x-3x-3x-3x	4x-4x-4x-4x-4x-4x	5x-5x-5x-5x-5x-5x-5x-5x	9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x-9x
Case 3	4x-4x-2x-2x	6x-6x-4x-4x-2x-2x	8x-8x-6x-6x 4x-4x-2x-2x	16x-16x-14x-14x-12x-12x-10x-10x-8x-8x-6x-6x-4x-4x-2x-2x

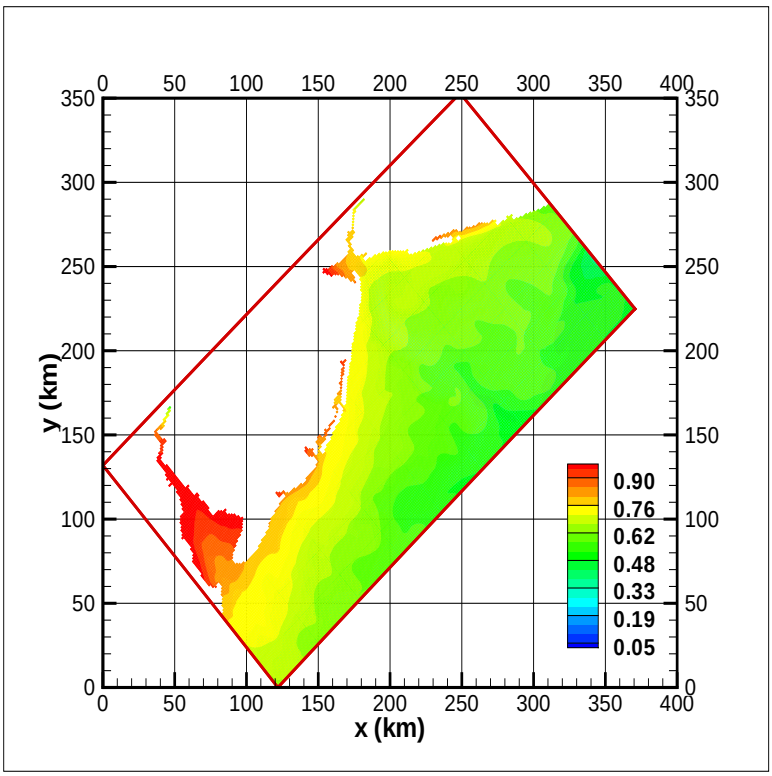

Fig. 1. New Jersey coast: Contours of the first POD mode (dimensionless temperature) and domain.

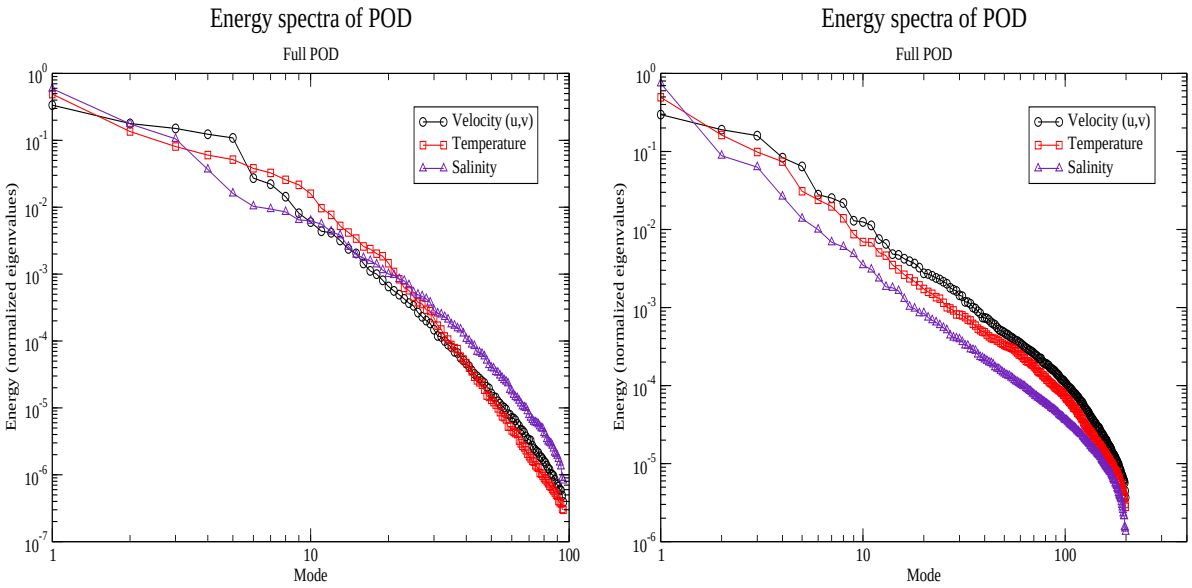

Fig. 2. New Jersey coast: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for 2.5-days (left) and 25-days (right) simulations.

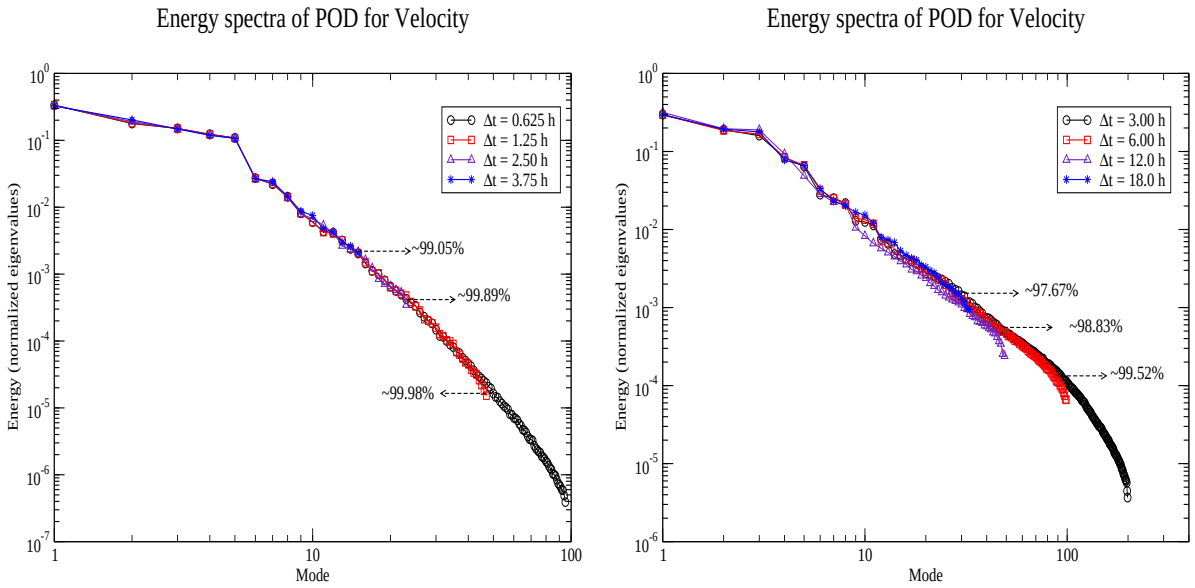

Fig. 3. New Jersey coast: Time resolution effects on the energy spectra of horizontal velocity (u, v) vector for 2.5-days (left) and 25-days (right) simulations.

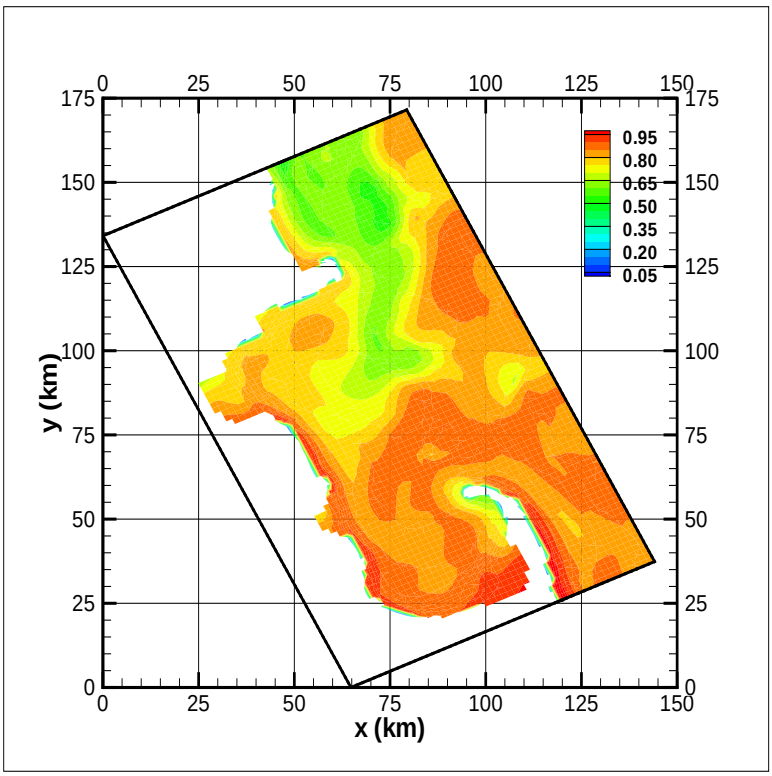

Fig. 4. Mass Bay: Contours of the first POD mode (dimensionless temperature) and domain.

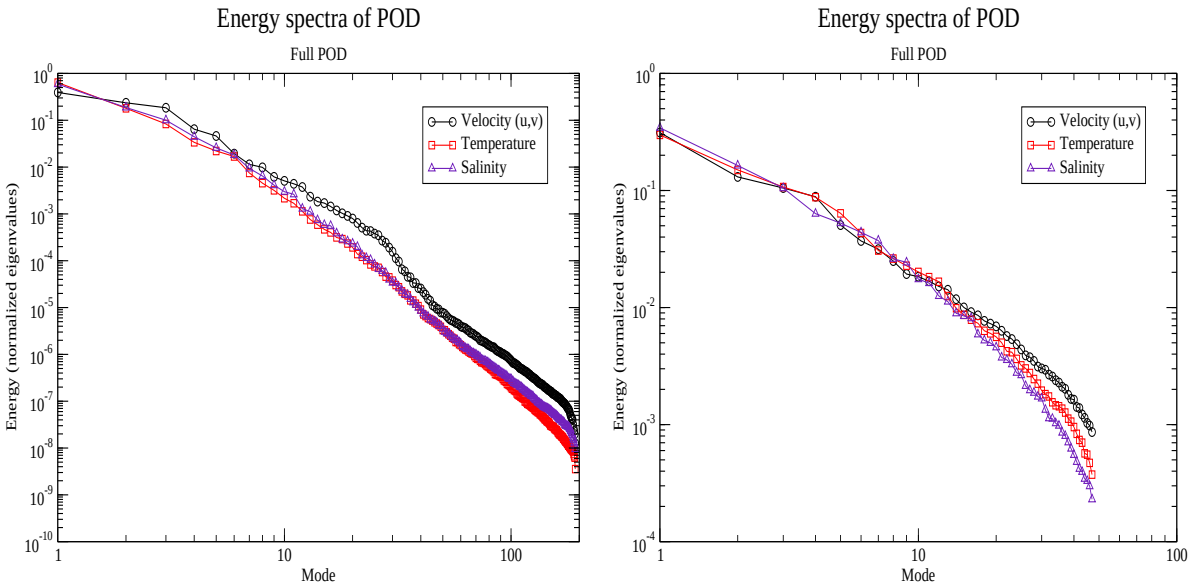

Fig. 5. Mass Bay: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for 8-days (left) and 47-days simulations.

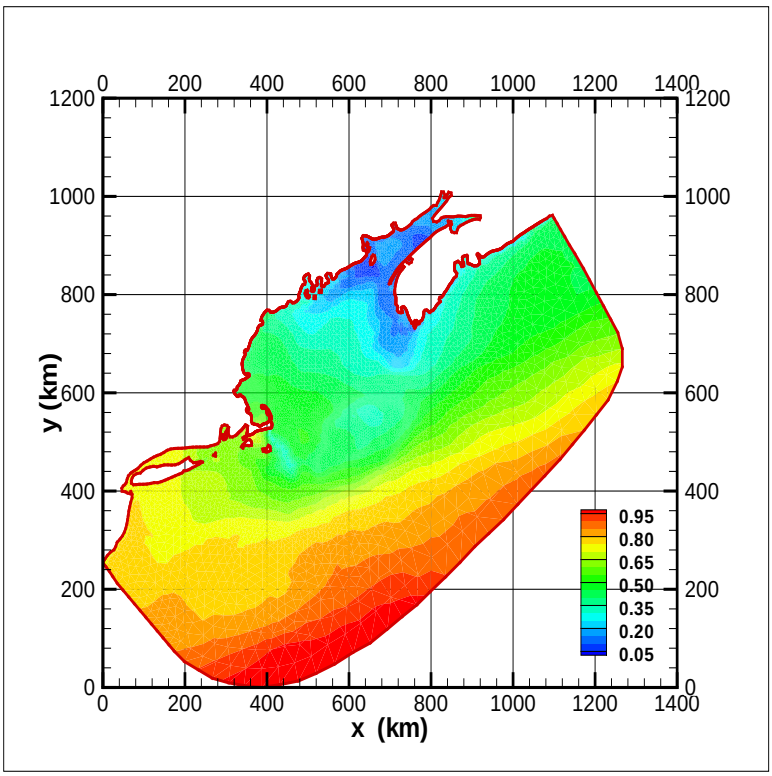

Fig. 6. Gulf of Maine: Contours of the first POD mode (temperature) and domain.

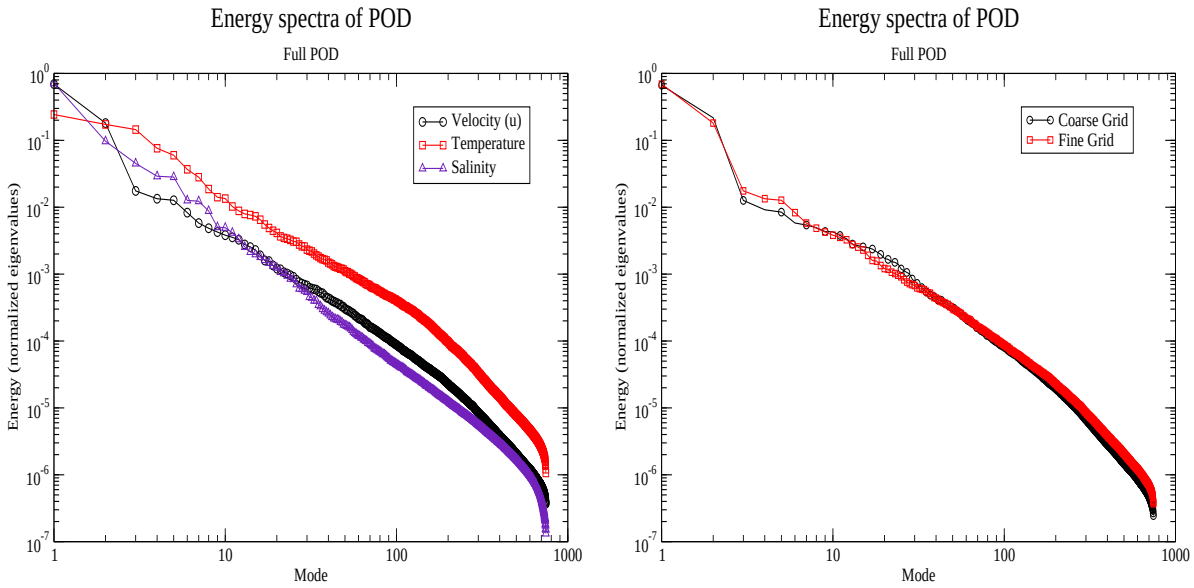

Fig. 7. Gulf of Maine. Left: Energy spectra of horizontal velocity vector (u, v), temperature and salinity for Gulf Of Maine. Right: Energy spectra of horizontal velocity vector at two different grid resolutions.

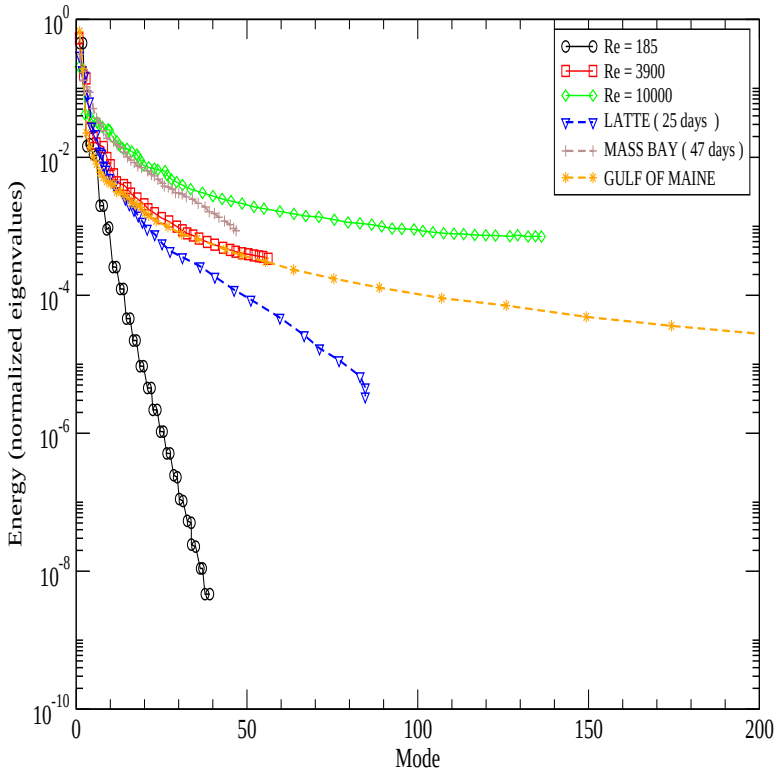

Fig. 8. Comparison of eigenspectra between the turbulent wake and the long-term ocean dynamics of three regions.

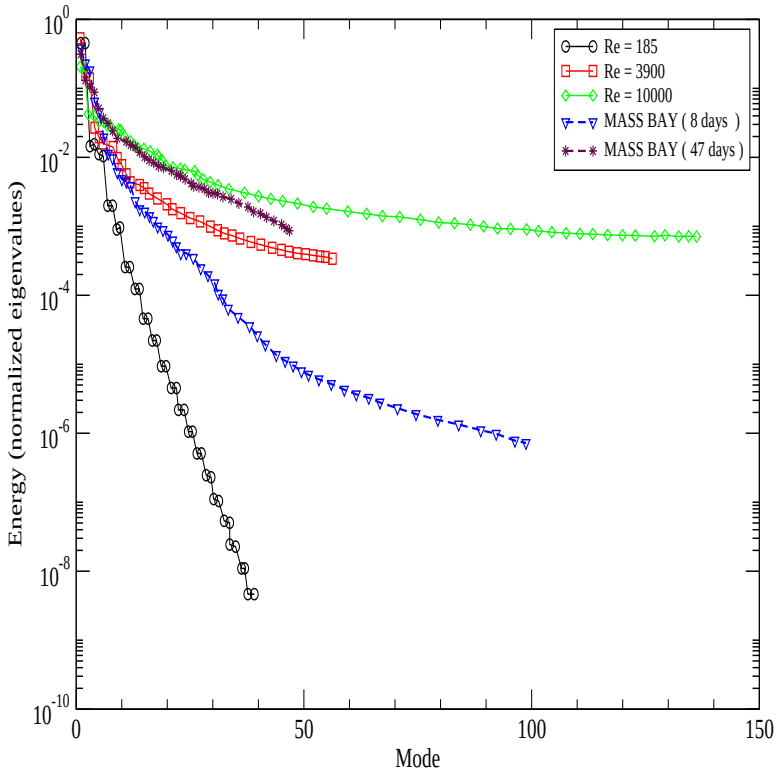

Fig. 9. Comparison of eigenspectra between the turbulent wake and the short-term and long-term ocean dynamics of Mass Bay.

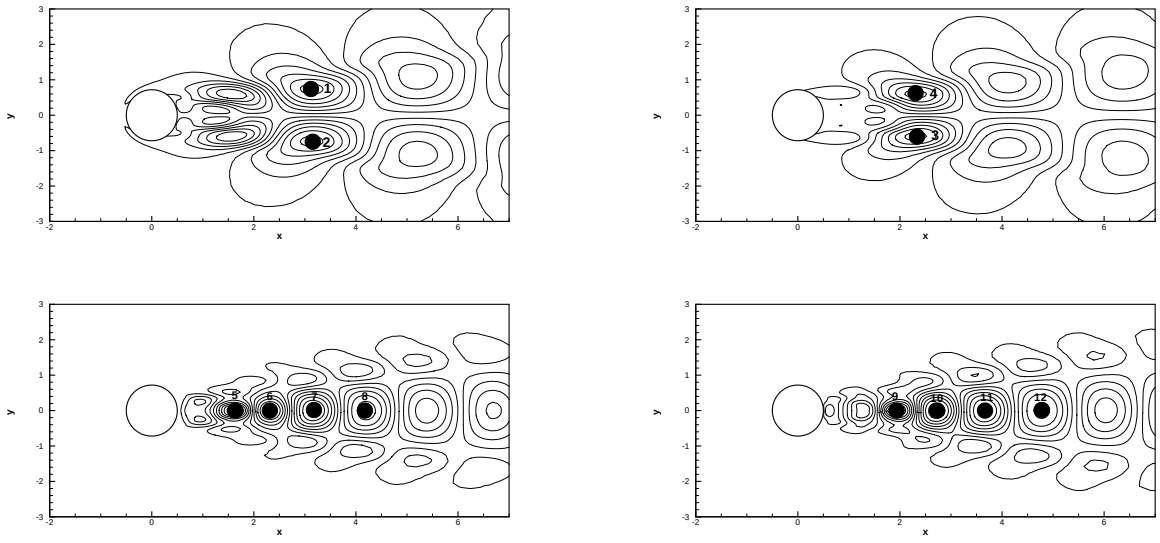

Fig. 10. Contours of first four POD modes of the total velocity U and sensor locations for case 1. Mode 1: upper-left; Mode 2: upper-right; Mode 3: lower-left; Mode 4: lower-right.

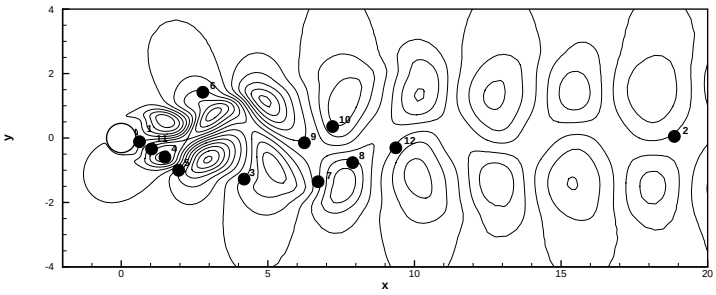

Fig. 11. Contours of streamwise velocity u and sensor locations for case 3.

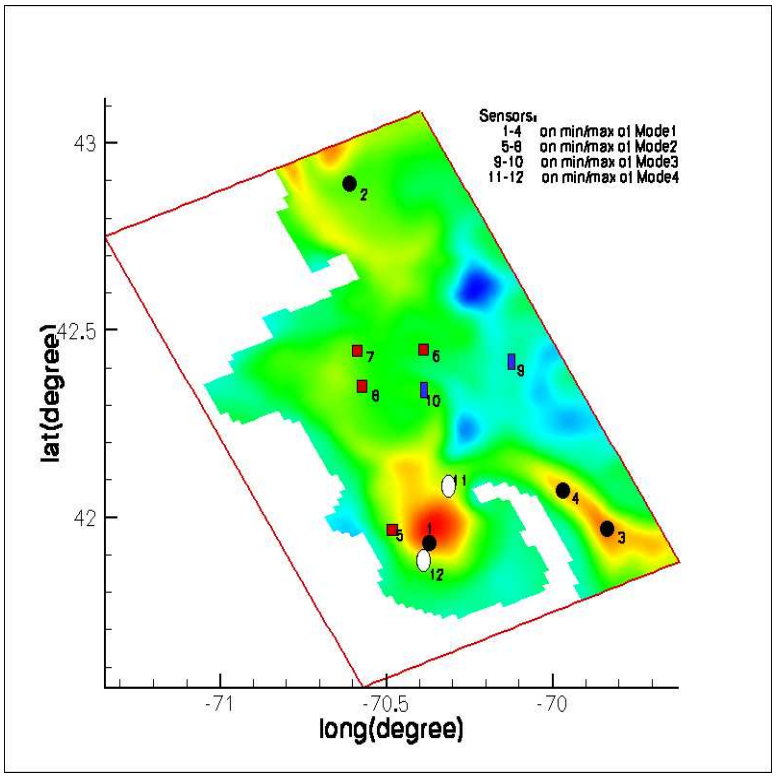

Fig. 12. Mass Bay: Schematic of best sensor locations for Case 1 (see Table 5: 4 modes, 12 sensors). Only a slice at the surface of the first POD mode is shown; different symbols correspond to different modes. The extrema are located at some depth from the surface. The contours represent the second POD mode of the temperature.

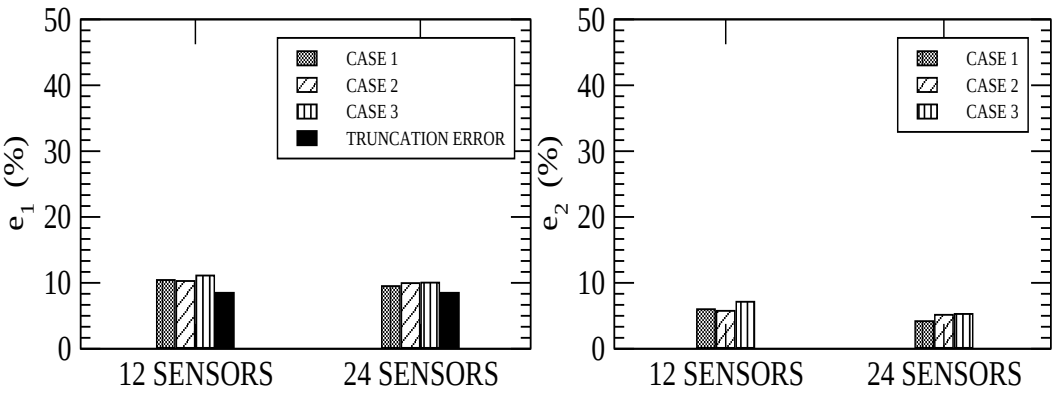

Fig. 13. Short-term simulation, 4-modes: Time averaged errors for total velocity. The total of *four* modes contain 92.6% of total energy and the corresponding truncation error is $\tilde{e}_1 = 8.5\%$. For 12 sensors $\kappa(M) = 14.82, 10.54, 9.55$ for Cases 1,2 and 3, respectively.

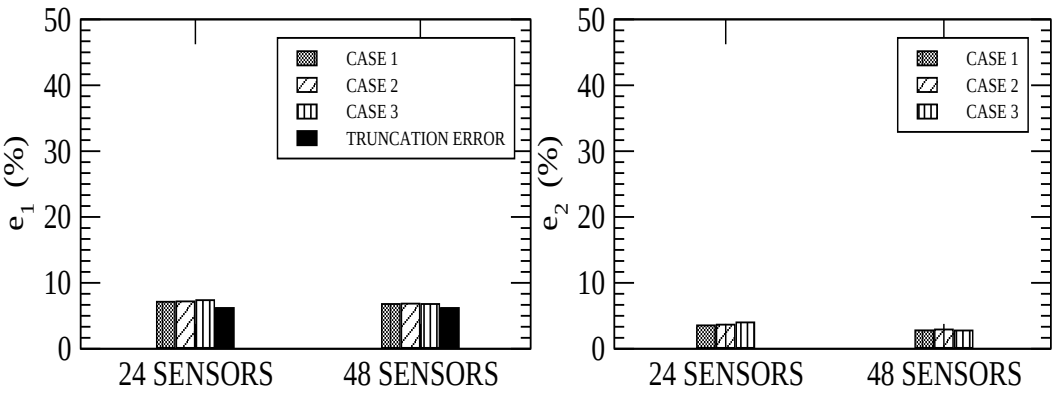

Fig. 14. Short-term simulation, 6-modes: Time averaged errors for total velocity. The total of *six* modes contain 96.1% of total energy and the corresponding truncation error is $\tilde{e}_1 = 6.2\%$. For 24 sensors $\kappa(M) = 18.92, 17.74, 16.54$ for Cases 1,2 and 3, respectively.

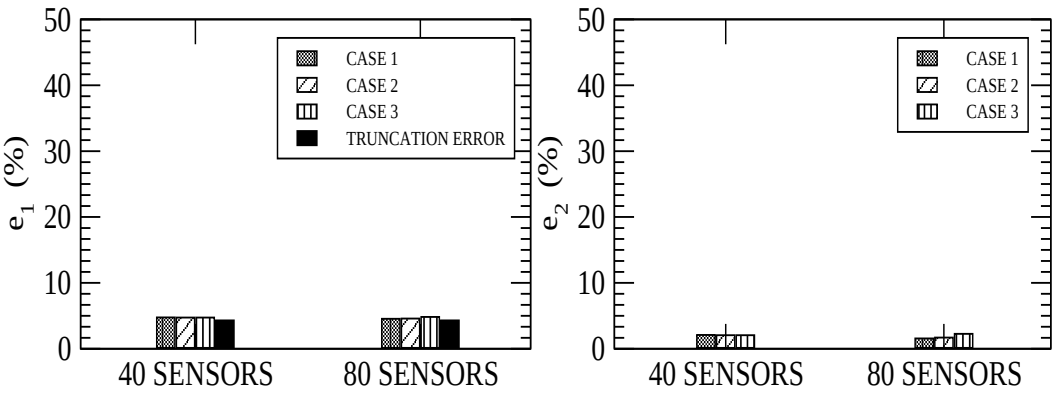

Fig. 15. Short-term simulation, 8-modes. Time averaged errors for total velocity. The total of *eight* modes contain 98.1% of total energy and the truncation error is $\tilde{e}_1 = 4.3\%$. For 40 sensors $\kappa(M) = 19.15, 20.37, 19.94$ for Cases 1,2 and 3, respectively.

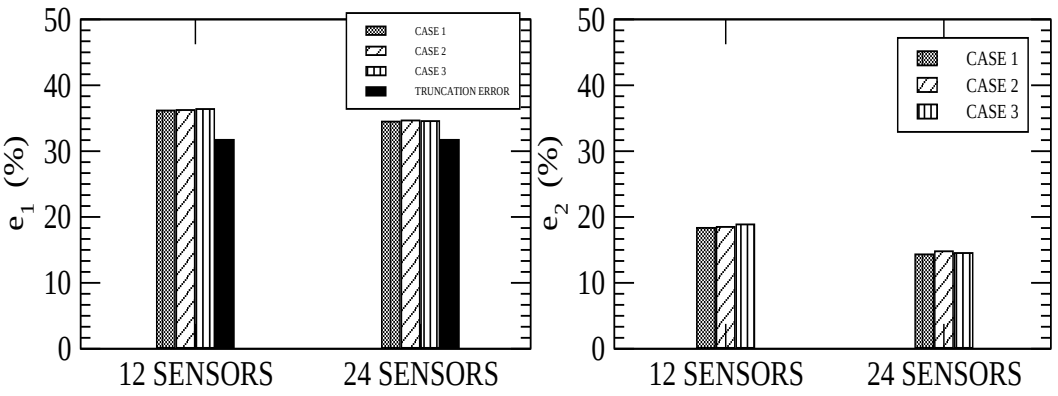

Fig. 16. Long-term simulation, 4-modes. Time averaged errors for total velocity. The total of *four* modes contain 53.2% of total energy and the truncation error is $\tilde{e}_1 = 31.7\%$. For 12 sensors $\kappa(M) = 20.73, 11.33, 7.10$ for Cases 1,2 and 3, respectively.

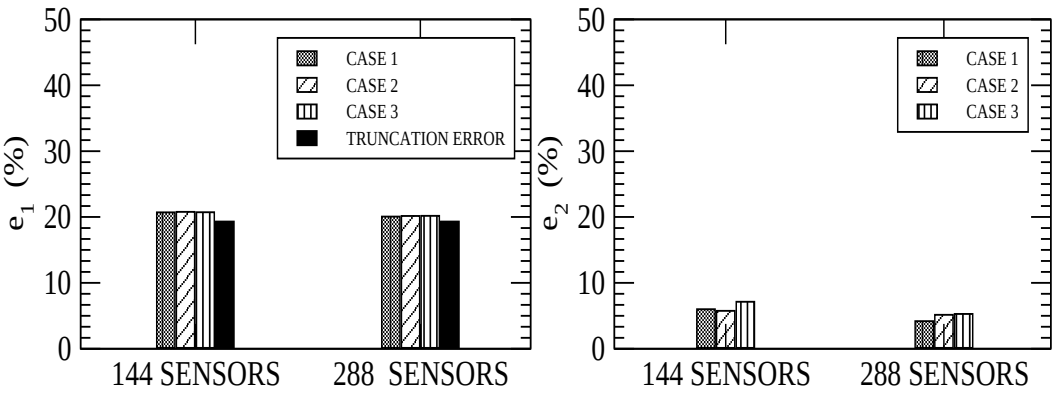

Fig. 17. Long-term simulation, 16-modes. Time averaged errors are for total velocity. The total of *sixteen* modes contain 83.1% of total energy and the truncation error is $\tilde{e}_1 = 19.3\%$. For 144 sensors $\kappa(M) = 13.42, 13.92, 11.78$ for Cases 1,2 and 3, respectively.

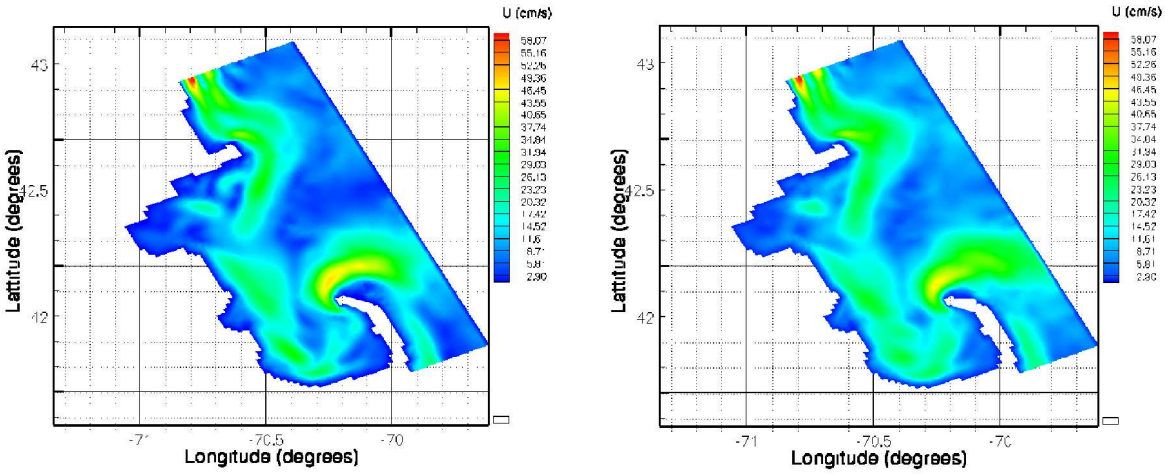

Fig. 18. Comparison between the original (left) and the reconstructed (right) total velocity field for the 24th snapshot in the long-term simulation.

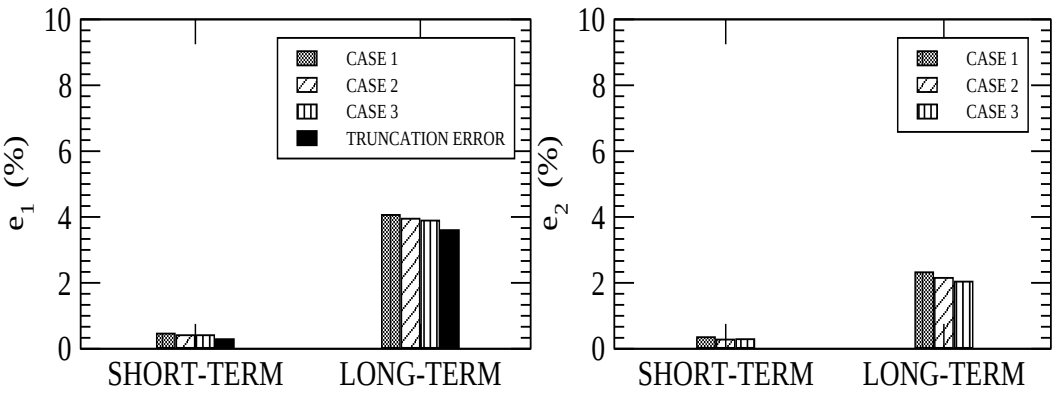

Fig. 19. Time averaged errors for temperature; 8-modes. Both short-term and long-term results are shown using 40 sensors.

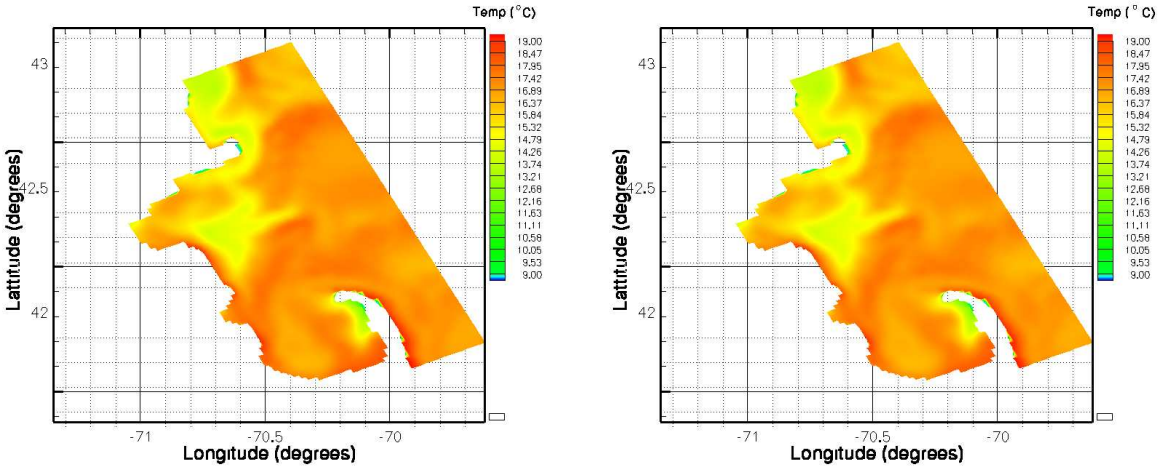

Fig. 20. Comparison between the original (left) and the reconstructed (right) temperature field for the 36th snapshot in the short-term simulation.