

Article

POD-Based Constrained Sensor Placement and Field Reconstruction from Noisy Wind Measurements: A Perturbation Study

Zhongqiang Zhang ¹, Xiu Yang ² and Guang Lin ^{3,*}

¹ Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, USA; zzhang7@wpi.edu

² Advanced Computing, Mathematics and Data Division, Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA; xiu.yang@pnnl.gov

³ Department of Mathematics & School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA

* Correspondence: guanglin@purdue.edu; Tel.: +1-765-494-1965

Academic Editor: Mohsen Zayernouri

Received: 12 January 2016; Accepted: 21 March 2016; Published: 14 April 2016

Abstract: It is shown in literature that sensor placement at the extrema of Proper Orthogonal Decomposition (POD) modes is efficient and leads to accurate reconstruction of the field of quantity of interest (velocity, pressure, salinity, *etc.*) from a limited number of measurements in the oceanography study. In this paper, we extend this approach of sensor placement and take into account measurement errors and detect possible malfunctioning sensors. We use the 24 hourly spatial wind field simulation data sets simulated using the Weather Research and Forecasting (WRF) model applied to the Maine Bay to evaluate the performances of our methods. Specifically, we use an exclusion disk strategy to distribute sensors when the extrema of POD modes are close. We demonstrate that this strategy can improve the accuracy of the reconstruction of the velocity field. It is also capable of reducing the standard deviation of the reconstruction from noisy measurements. Moreover, by a cross-validation technique, we successfully locate the malfunctioning sensors.

Keywords: proper orthogonal decomposition; sensor placement; uncertainty; anomaly detection

1. Introduction

Recently, a great effort has been carried out by scientists in developing efficient methods for sensor placement and ocean state reconstruction based on limited measurements that can be used to improve ocean forecasting, at least for regional modelling. One of the efficient methods is the Proper Orthogonal Decomposition (POD), see [1–5], also known as the method of Empirical Orthogonal Functions (EOFs), which have been applied to analyze the measurement data and to develop reconstruction procedures for gappy data sets, see [6–16].

The success of POD approaches depends on a low-dimensional representation of the ocean processes, given the wide 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 s to several months ([17]). In [18], the authors showed that only a few spatio-temporal POD modes are sufficient to describe the most energetic ocean dynamics. In particular, they demonstrated that the extrema of the POD spatial modes are very good locations for sensor placement and accurate field reconstruction given a limited number of observing stations and assuming perfect measurements. In [19], the authors proposed to place sensors at extrema but away from each other within a given distance as they found

that the extrema of POD spatial modes can be very close and thus produce redundant measurements. See also [20] for a similar approach.

The objective of this paper is to extend the method developed in [19] by investigating the effect of noisy/bad measurements as well as detecting malfunctioning sensors, when POD basis is known. As in [19], we will also impose “declustering” constraints on the sensors as we have observed to avoid redundant measurements at near-by points. Because we use the extrema of the first few modes, empirically, the condition number of the matrix in the linear system of gappy pod method is reduced as shown in [18,19]. We find that this strategy also helps to reduce the reconstruction error. We will consider realistic scenarios, such as measurement uncertainty and bad data from malfunctioning sensor(s). In order to demonstrate the new ideas, we will test our method with the 24 hourly spatial wind field simulation data sets for the Maine Bay. For simplicity, we assume that the noisy measurement data is perturbed by some Gaussian processes. When the perturbation is at the magnitude of the measurements, we assume that the sensor is malfunctioning and thus will be excluded from reconstruction. With a heuristic cross-validation technique, we can locate the malfunctioning sensor effectively.

We remark that there are many other ways to accommodate noise into models and predict the state space from reconstruction—for example, Kalman filtering [21], empirical interpolation method [22,23], and proper generalized decomposition [24,25]. The purpose of our paper is to understand the effect of noisy and bad measurements within the POD framework. Here, we first obtain POD modes from exact complete field and use these POD modes to perform a least square fitting of limited data with the noise in the measurement. This procedure of reconstructing missing or gappy data with POD basis was developed by Everson and Sirovich [26]. While not discussed here, we also note that if the original snapshot ensemble has incomplete data, the POD basis vectors can be computed using an iterative gappy approach [19,27,28]. For bad measurements, we do not perform a Gappy-POD analysis. Instead, we still use the POD modes obtained from the exact field. We alter some measurements of a certain sensor which is regarded as malfunctioning. This is definitely an ideal case for the POD analysis since the bad measurements may significantly change the POD basis. Nevertheless, this ideal testbed provides a basic understanding of the robustness of POD analysis for functioning sensor measurements.

The rest of the paper is organized as follows. In Section 2, we first review the sensor placement strategy based on POD and introduce how to apply an exclusion disk approach to avoid redundant measurements. We discuss in Section 3 the effects of noisy measurements and the effects of exclusion disk approach. In Section 4, we show how to detect a malfunctioning sensor before we summarize and end the paper.

2. A POD-Based Sensor Placement Strategy

The key idea of the POD method is to expand a state variable $u(x, t)$ at certain time t in a spatial domain X as an orthogonal series

$$u(x, t) = \sum_{k=1}^{\infty} a_k(t) \Phi_k(x), \quad (1)$$

where $\Phi_k(x)$'s are orthonormal spatial modes in $L^2(X)$. Expansion (1) is normally truncated after K terms, which represents the dimensionality of the system. Then, number K is usually small so that we have a low dimensional representation of the process $u(x, t)$ while $\|u(\cdot, t) - \sum_{k=1}^K a_k(t) \Phi_k\| \leq \epsilon \|u(\cdot, t)\|$ and ϵ is small, say $\epsilon = 0.1$.

In this paper, we will use POD modes obtained from the exact complete field and hence they are “perfect” POD modes. We first review the Gappy-POD before we discuss our sensor placement strategy and the effect of noisy measurements.

2.1. A Review of Gappy-POD

In practice, $u(x, t)$ may be only obtained at limited locations. To deal with such data (called gappy data), a so-called Gappy-POD was proposed in [7] in order to reconstruct gappy flow fields from a limited number of measurements. As a brief review, we adopt the introduction in [19,27]. Consider a scalar gappy field $\tilde{u}(x, t)$ as a point-wise product of an indicator function $m(x, t)$ and a complete field $u(x, t)$, i.e.,

$$\tilde{u}(x, t) \stackrel{\text{def}}{=} m(x, t) u(x, t). \quad (2)$$

The indicator function $m(x, t)$ has values 0 or 1 depending on whether we have data at the corresponding space-time location or not. The goal is to construct a reliable estimator $\hat{u}(x, t)$ of $u(x, t)$ in the space-time regions where $m(x, t) = 0$. In the Gappy-POD framework, it is done applying the method of least-squares: we look for a representation of $\hat{u}$ in the form

$$\hat{u}(x, t) = \sum_{k=1}^K b_k(t) \Phi_k(x), \quad (3)$$

where Φ_k are normalized POD spatial modes of $\hat{u}$ and b_k are unknown coefficients. Note that in practice, Φ_k 's, are extracted from simulation or data assimilation results, and in this paper they are from the perfect model (1). Then, for each snapshot (e.g., each time t), we minimize the approximation error between u and $\hat{u}$ in the gappy norm

$$\|u - \hat{u}\|_m^2 \stackrel{\text{def}}{=} (u - \hat{u}, u - \hat{u})_m = (u, u)_m - 2 \sum_{k=1}^K b_k (u, \Phi_k)_m + \sum_{k=1}^K \sum_{j=1}^K b_k b_j (\Phi_k, \Phi_j)_m, \quad (4)$$

where the (gappy) inner product $(\cdot, \cdot)_m$ is defined as

$$(a, b)_m \stackrel{\text{def}}{=} \int_X m(x, t) a(x) b(x) dX, \quad \forall a, b \in L^2(X). \quad (5)$$

Minimizing Equation (4) with respect to b_k yields the linear system

$$\mathbf{M} \mathbf{b} = \mathbf{f}, \quad (6)$$

where $M_{kj} \stackrel{\text{def}}{=} (\Phi_k, \Phi_j)_m$ and $f_j \stackrel{\text{def}}{=} (u, \Phi_j)_m$. Then, we can obtain time-coefficient vector $\mathbf{b} = [b_1(t), \dots, b_K(t)]$ from Equation (6).

Remark 1. One should be careful to solve Equation (6) efficiently as the condition number of $\mathbf{M}$ varies with $m(x, t)$. The matrix $\mathbf{M}$ reduces to a $K \times K$ identity matrix in case of complete data, i.e., $m(x, t) = 1$. For gappy data, the condition number of $\mathbf{M}$ can be small or large depending on the $m(x, t)$. When the condition number of $\mathbf{M}$ is large, we still formally write Equation (6), but we will solve its corresponding overdetermined system rather than Equation (6). If $\mathbf{M}$ is singular, we seek the pseudoinverse. From now on, we keep this notation and will not state it explicitly.

Based on the aforementioned POD modes and the reconstruction procedure, the authors in [27,29] demonstrated that effective sensor placement strategies can be designed. In this case, the indicator function $m(x, t)$ is defined through sensor locations, i.e., $m(l_j, t) = 1$ if there is a (working) sensor at position l_j —zero otherwise. The only difference is that the complete field u is now substituted by sensor measurements $d(l_j, t)$ ($j = 1, \dots, N_s$), where l_j denotes the position of the j -th sensor and N_s is the total number of sensors. The resulting linear system will be also denoted by Equation (6).

In this work, we use the 24 hourly spatial wind field simulation data sets in Maine Bay simulated from the Weather Research and Forecasting (WRF) model. The data we use contain 24 snapshots of the wind field within an area of about $35 \text{ km} \times 25 \text{ km}$ from 13 July 2004 00:00 a.m. to 14 July 2004 00:00 a.m. The computation domain and grid are illustrated in Figure 1. The time

difference between two successive snapshots is 1 h, namely, snapshot 1 corresponds to data on 13 July 2004 at 0:00 a.m., snapshot 2 corresponds to the data on 13 July 2004 at 1:00 a.m., *etc.* The temporal average of the window field is subtracted from the field, hence we concentrate on the fluctuation. By cross-correlating different snapshots obtained from the wind simulations, we construct the covariance matrix and its eigen-decomposition. From the eigen-decomposition, we obtain the POD eigenvalues and corresponding hierarchical modes. Here, we use Equations (2)–(6) to obtain Gappy-POD modes. We note that in [27], the procedure Equations (2)–(6) is iterated when the measurement data is incomplete. Here, we focus on complete data but with noise.

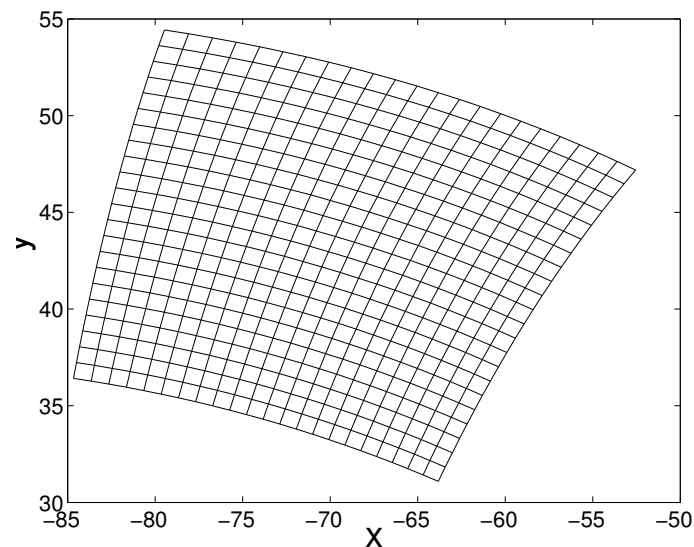


Figure 1. Computation domain and discretization mesh.

In Figure 2, we show the normalized POD eigenvalue spectra of total velocity. For illustration, in Figure 3, we show the contour plots of the first and second POD spatial modes for total velocity.

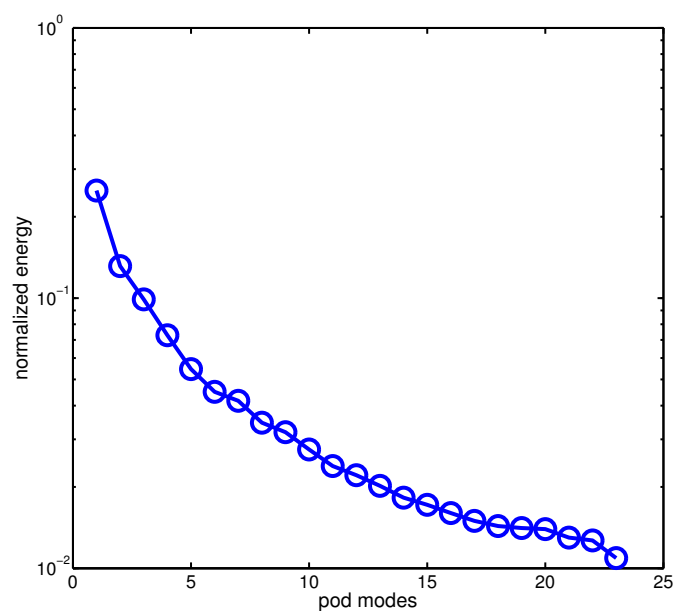


Figure 2. Normalized POD spectra of total velocity.

In case of complete data, e.g., $m(x, t) = 1$, the L_2 error of the truncated approximation for each snapshot in Equation (4) is:

$$\epsilon_c \stackrel{\text{def}}{=} \|u - \hat{u}\|^2 = \left\| \sum_{k=1}^N a_k \Phi_k - \sum_{k=1}^K a_k \Phi_k \right\|^2 = \left\| \sum_{k=K+1}^N a_k \Phi_k \right\|^2 = \sum_{k=K+1}^N a_k^2, \quad (7)$$

due to the orthogonality of Φ_k and N is the total number of snapshots, i.e., $N = 24$. We note that the ϵ_c is the lower bound of the error by Gappy-POD method. Since we use gappy data to reconstruct the entire field, the coefficients for the POD models may not be optimal, the coefficients b_k 's in Equation (3) are not equal to a_k 's in Equation (1). Therefore, the reconstruction error for a fixed snapshot (fixed t) is:

$$\epsilon_g \stackrel{\text{def}}{=} \left\| \sum_{k=1}^N a_k \Phi_k - \sum_{k=1}^K b_k \Phi_k \right\|^2 = \left\| \sum_{k=1}^K a_k \Phi_k - \sum_{k=1}^K b_k \Phi_k \right\|^2 + \left\| \sum_{k=K+1}^N a_k \Phi_k \right\|^2 = \sum_{k=1}^K (a_k - b_k)^2 + \sum_{k=K+1}^N a_k^2. \quad (8)$$

In [18,19], the authors demonstrate numerically the convergence of the reconstruction error of the Gappy-POD method: ϵ_g becomes closer to ϵ_c as the number of sensors increases. More specifically, $\sum_{k=1}^K (a_k - b_k)^2 \rightarrow 0$ as number of sensors increases. In our data, the first six POD modes ($K = 6$) capture more than 90% percent of the total energy. Throughout the paper, we set $K = 6$ and we select locations for sensors based on these six modes.

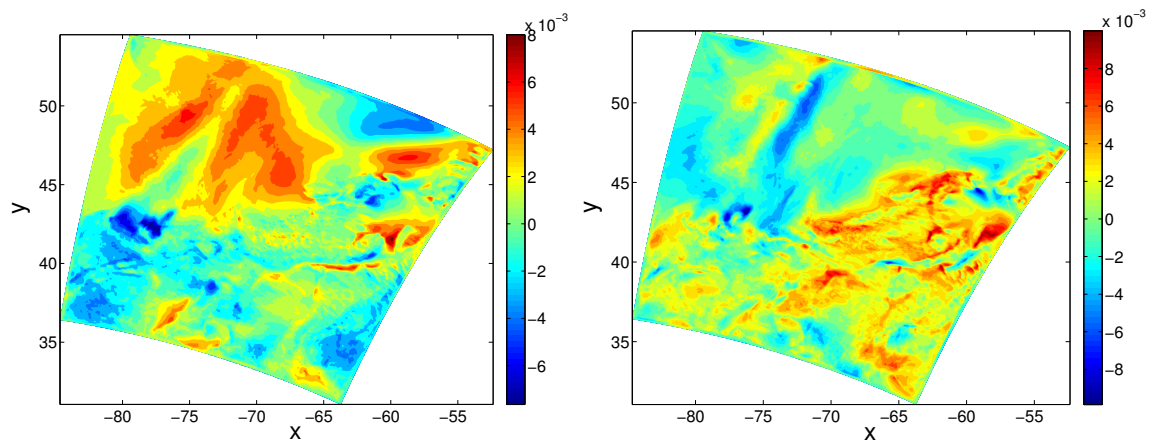


Figure 3. Contour plots of the first (left) and second (right) POD mode.

2.2. Constrained Sensor Placement

We first consider the problem of finding the best possible locations where to deploy a limited number of sensors to reconstruct the entire field with existing *exact* POD modes. Authors in [18,30] demonstrate that the extrema of POD modes are very good locations, when examining reconstruction error in l_2 norm, see Equations (10) and (11), to place the sensors for regional ocean forecasting. Yang *et al.* [19] further introduced the exclusion cylinder to further increase the accuracy of the reconstruction by reducing the redundancy as locations of extrema of different POD modes can be very close. As shown in [19], the exclusion cylinder also helps to reduce the variance of the reconstructed field if the measurement is polluted by the noise. In this paper we employ the exclusion disk instead of the exclusion cylinder for a 2D model. Specially, we impose the following constraint that the distances between each two sensors are larger than R . That is, if we place two sensors at the point (x_1, y_1) and another point (x_2, y_2) , then the following restraint is imposed:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} > R. \quad (9)$$

In our wind model, the radius R will be expressed in kilometers. This is achieved by a greedy algorithm: we start with a empty set $F = \emptyset$, then for the next candidate location (x^*, y^*) of the sensor (an extrema of a specific mode), we examine whether $\sqrt{(x^* - x_i)^2 + (y^* - y_i)^2} > R$ holds for each point (x_i, y_i) in F . If not, we neglect this location and examine the next candidate. When a qualified new sensor location $(\tilde{x}, \tilde{y})$ is found, we expand F as $F \leftarrow F \cup (\tilde{x}, \tilde{y})$.

We distribute the sensors evenly according to the extrema of the POD modes we use to reconstruct the field. In Table 1, we provide examples of sensor distribution for a fixed number of sensors with different numbers of POD modes reconstructing the field. With a configuration denoted as “ $2s$ - $2s$ - $2s$ - $2s$ ” with $s = 1$, we distribute 2, 2, 2 and 2 sensors associated with the POD modes 1, 2, 3 and 4, respectively. Similarly a configuration denoted as “ $4s$ - $4s$ - $4s$ - $4s$ - $4s$ - $4s$ ” with $s = 2$ means that we use eight sensors in connection with each POD mode from 1 to 6. Here, we employ this even distribution to demonstrate our idea. The numerical results in [18,19] demonstrate that different configuration of sensor placement will not induce significant difference in the reconstruction.

Table 1. Sensor network configurations in modal space with a fixed number of sensors. A configuration denoted as “ $4s$ - $4s$ - $4s$ - $4s$ - $4s$ - $4s$ ” with $s = 2$ means that we are using 8 sensors in connection with each POD mode from 1 to 6.

Four Modes	Six Modes	Eight Modes
6s-6s-6s-6s	4s-4s-4s-4s-4s-4s	3s-3s-3s-3s-3s-3s-3s-3s

In our numerical tests, we fix the number of sensors to 24 and use six modes (hence the configuration is “ 4 - 4 - 4 - 4 - 4 - 4 ”) to reconstruct the entire field to see the effect of the exclusion disk size in reconstructing the total velocity. Figure 4 presents the locations of the sensors with different sizes of the exclusion disk. The contour is for the first POD mode. With larger R , the radius of the disk, the sensors are distributed further away from each other.

Remark 2. To avoid redundant measurements, the authors [25] suggested an approach without using the exclusion disk strategy. However, the approach therein only allows one sensor per mode.

To determine the effect of the different exclusion disk size, we define the reconstruction error as (for each flow snapshot)

$$e_i^2 \stackrel{\text{def}}{=} \frac{\int_X (\hat{U}_i - U_i)^2 dX}{\int_X (U_i)^2 dX}, \quad (10)$$

where U_i is the measurement of total velocity at the i -th snapshot and $\hat{U}_i$ denotes the estimator of the i -th snapshot obtained by solving Equation (6). Here, the integration is performed on the computational domain in physical space. To measure the reconstruction error within the whole period of interest, we also define the time-averaged error as

$$e_{avg}^2 \stackrel{\text{def}}{=} \frac{1}{S} \sum_{i=1}^S e_i^2, \quad (11)$$

where S denotes the total number of available snapshots within the considered time interval. The time-averaged error for $R = 0, 0.5, 1, 2$ are presented in Table 2. Similar to the cases studied in [19] a proper selection of the size of exclusion disk helps to improve the accuracy of the reconstruction. In addition, the size of the exclusion disk need not be very large as we can observe that the error for $R = 0.5, 1, 2$ are quite close. In addition, the condition number of matrix M in Equation (6) decreases as R increases, which is consistent with the results reported in [18,19]. As an illustration, we compare the exact velocity field of the first snapshot (13 July 2004 00:00 a.m.) with those reconstructed with $R = 0$ and $R = 2$, respectively in Figure 5. It is demonstrated in Figure 5 that the reconstruction of the wind field at the first snapshot is more accurate with the exclusion disk than that without the exclusion disk.

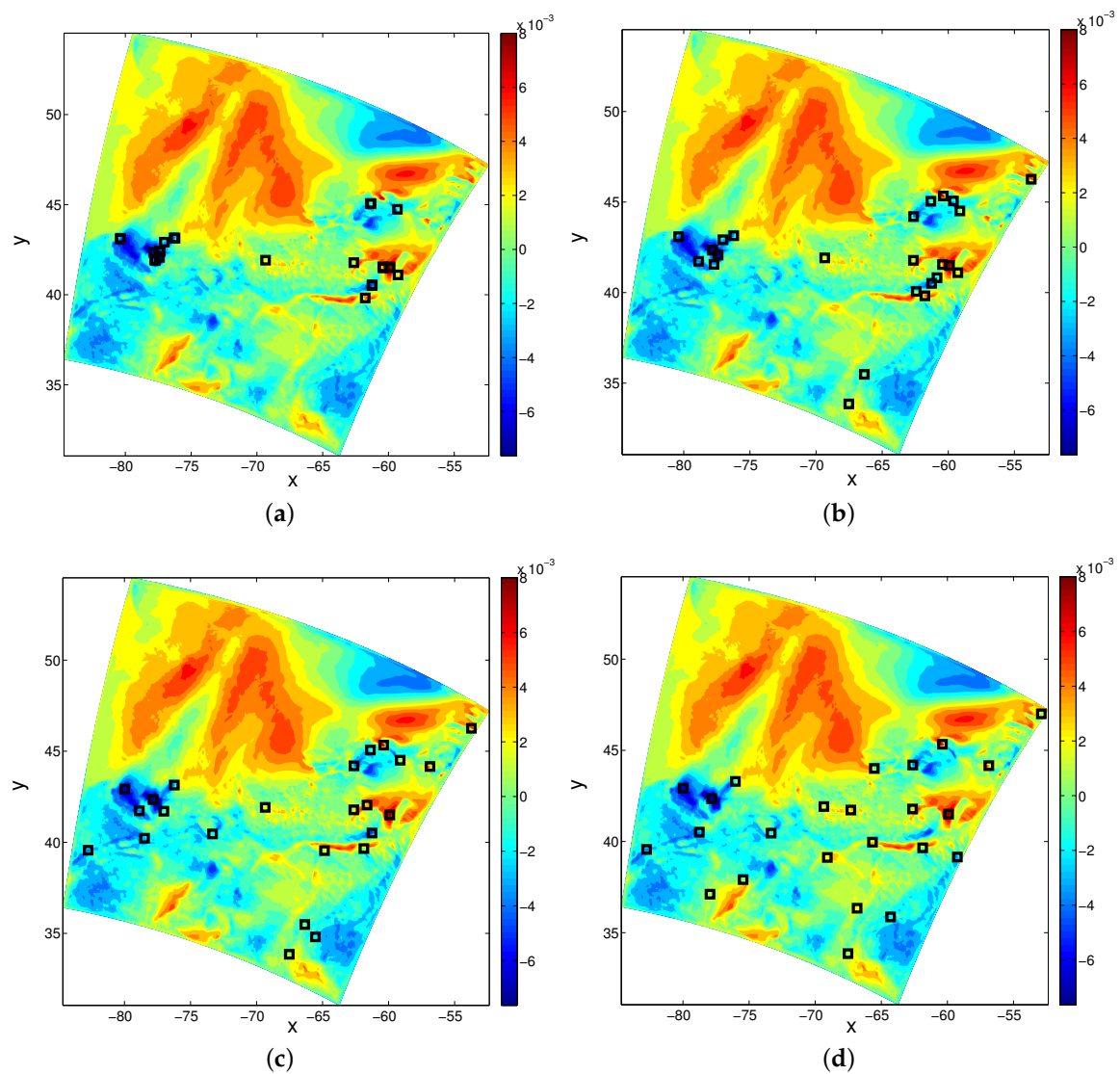


Figure 4. Location of 24 sensors placed on the extrema of the first six modes superimposed on the contour of the first mode with different size of the exclusion disk. (a) $R = 0$; (b) $R = 0.5$; (c) $R = 1$; (d) $R = 2$.

Table 2. e_{avg}^2 and condition number of M ($\text{Cond}(M)$) for different exclusion disk size R .

R Value	$R = 0$	$R = 0.5$	$R = 1$	$R = 2$
e_{avg}^2	17.27%	10.11%	9.26%	9.02%
$\text{Cond}(M)$	83.34	22.01	16.45	10.10

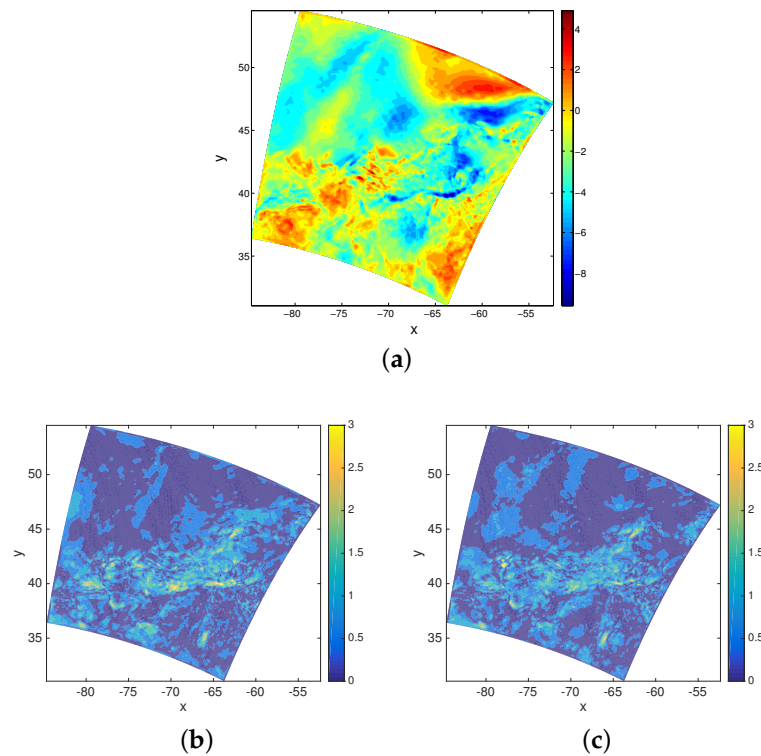


Figure 5. Comparison of the reconstructed field of the first snapshot with different sizes of exclusion disks. (a) exact; (b) $R = 0$, absolute errors; (c) $R = 2$, absolute errors.

3. Uncertainty in Measurements

In this section, we investigate the uncertainty of the reconstructed field assuming that the sensor signals are perturbed by random noise. Specifically, we assume that the total velocity detected by the sensors has the form

$$d(l_i, t; \xi) = U(l_i, t) + \xi(l_i, t), \quad (12)$$

where $U(l_i, t)$ is a total velocity to be evaluated and $\xi(l_i, t)$ are N_s i.i.d. zero-mean Gaussian processes with standard deviation $\sigma_\xi(l_i, t)$. To accommodate the measurement uncertainty, we look for an estimation of the total velocity by random temporal modes and deterministic spatial modes [19,31,32] as:

$$U_{\text{est}}(x, t_j; \xi) = \sum_{k=1}^K \eta_k(t_j; \xi) \Phi_k(x). \quad (13)$$

Here, Φ_k are obtained by decomposing full wind field perfect data from Equation (3). The notation $U_{\text{est}}(x, t_j; \xi)$ emphasizes that the stochastic total velocity Equation (13) depends on all random variables

$$\xi \stackrel{\text{def}}{=} [\xi(l_1, t_1), \dots, \xi(l_{N_s}, t_s)], \quad (14)$$

characterizing the measurement errors. Now, let us consider the following functional (more general data assimilation schemes based on quadratic *regularization functionals* can be considered).

$$\mathcal{J}[\eta_k] = \left\langle \sum_{i=1}^{N_s} [d(l_i, t; \xi) - \sum_{j=1}^K \eta_j(t; \xi) \Phi_j(l_i)]^2 \right\rangle, \quad (15)$$

where $\langle \cdot \rangle$ denotes an average with respect to the joint probability density of the random vector ξ . Minimization of Equation (15) with respect to $\eta_k(t; \xi)$ yields the following Euler-Lagrange equations

$$\sum_{k=1}^K \eta_k(t; \xi) \sum_{i=1}^{N_s} \Phi_k(l_i) \Phi_m(l_i) = \sum_{i=1}^{N_s} d(l_i, t; \xi) \Phi_m(l_i). \quad (16)$$

This system can be recast in same form as Equation (6), i.e.,

$$\sum_{j=1}^K (\Phi_i, \Phi_j)_m \eta_j = (d, \Phi_i)_m, \quad (17)$$

provided the gappy inner product Equation (5) is defined in terms of the sensor locations; that is, $m(x, t) = 1$ if $x = l_j$ —zero otherwise. Denote by $\mathbf{M}^{-1}$ the (pseudo)inverse of $\mathbf{M}$, then we obtain a unique solution to Equation (17) in the form

$$\eta_i(t; \xi) = \sum_{j=1}^K (\mathbf{M}^{-1})_{ij} (d, \Phi_j)_m. \quad (18)$$

Next, we calculate the standard deviation of the estimate Equation (13) based statistical assumption of the measurements errors. Specifically, by using the independence hypothesis of the measurement errors, we obtain

$$\sigma_{U_{\text{est}}}(x, t) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^K \sum_{k=1}^K (\langle \eta_i - \langle \eta_i \rangle \rangle \langle \eta_k - \langle \eta_k \rangle \rangle) \Phi_i(x) \Phi_k(x)}, \quad (19)$$

where we have from Equations (6), (18) and (12) that

$$\langle \eta_i \rangle = \sum_{j=1}^K (\mathbf{M}^{-1})_{ij} (\langle d \rangle, \Phi_j)_m, \quad \eta_i - \langle \eta_i \rangle = \sum_{j=1}^K (\mathbf{M}^{-1})_{ij} (\xi(l_j, t), \Phi_j(x))_m. \quad (20)$$

Substituting Equations (20) into (19) yields the following final formula

$$\sigma_{U_{\text{est}}}(x, t) = \sqrt{\sum_{i=1}^K \sum_{k=1}^K \sum_{l=1}^K \sum_{n=1}^K \sum_{j=1}^{N_s} \sigma_{\xi}^2(l_j, t) (\mathbf{M}^{-1})_{il} (\mathbf{M}^{-1})_{kn} \Phi_l(l_j) \Phi_n(l_j) \Phi_i(x) \Phi_k(x)}. \quad (21)$$

3.1. Results for Uniform Measurement Errors

Let us assume that the standard deviation of the measurement errors is a constant, i.e., $\sigma_{\xi}(x, t) = \sigma_{\xi}$. In this case, Equation (21) simplifies to

$$\sigma_{U_{\text{est}}}(x) = \sigma_{\xi} \sqrt{\sum_{j=1}^{N_s} \left(\sum_{n=1}^K \sum_{k=1}^K (\mathbf{M}^{-1})_{kn} \Phi_n(l_j) \Phi_k(x) \right)^2}, \quad (22)$$

yielding to a time-independent standard deviation of the estimate Equation (13).

Figure 6 reports on results of standard deviation calculations given $\sigma_{\xi} = 1.0$ for different sensor placement strategies. In particular, we fix the POD modes used to reconstruct the field and compare the standard deviation of the reconstructed field with different sizes of exclusion disks. The results of reconstructing the first snapshot are presented. As the reconstructions of all snapshots are similar, we will not present the corresponding results here.

In Table 3, we compare the averages and the maximum of $\sigma_{U_{\text{est}}}(\mathbf{x})$ over all $\mathbf{x}$. Numerical results indicate that the exclusion disk method indeed yields smaller standard deviations of the reconstructed field. Clearly, a suitable R helps to reduce the standard deviation Equation (22) in the prediction.

Table 3. Average and maximum of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$ over all snapshots given $\sigma_{\xi} = 1$ for different R .

R Value	$R = 0$	$R = 0.5$	$R = 1$	$R = 2$
maximum of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$	1.383	0.9868	0.9305	0.6744
average of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$	0.1781	0.1104	0.1100	0.1031

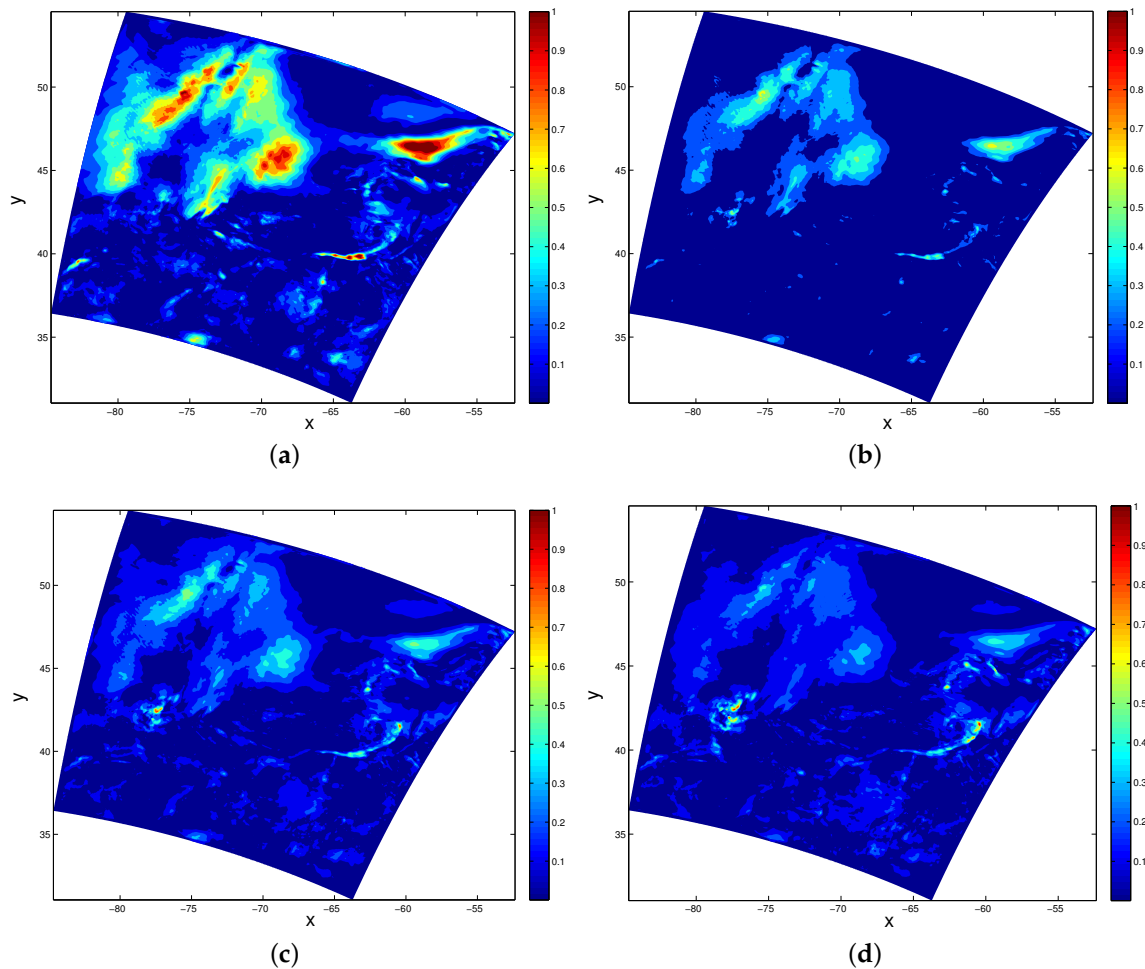


Figure 6. Standard deviation of the reconstructed snapshot 1 based on noisy measurement which does not depends on the wind field. (a) $R = 0$; (b) $R = 0.5$; (c) $R = 1$; (d) $R = 2$.

3.2. Results for Non-Uniform Measurement Errors

Next, we study a more realistic case in which the standard deviation of $\xi(\mathbf{x}, t)$ is functionally dependent on the total velocity $U(\mathbf{x}, t)$. The simplest case is $\sigma_{\xi}(\mathbf{x}, t) = \alpha U(\mathbf{x}, t)$, where α is a positive constant ($U \geq 0$ by construction). From Equation (21), we readily derive the following analytical expression for the standard deviation of U_{est}

$$\sigma_{U_{\text{est}}}(\mathbf{x}, t) = \alpha \sqrt{\sum_{j=1}^{N_s} U^2(I_j, t) \left(\sum_{n=1}^K \sum_{k=1}^K (M^{-1})_{kn} \Phi_n(I_j) \Phi_k(\mathbf{x}) \right)^2}. \quad (23)$$

Here, we obtain numerical results with a typical value $\alpha = 1.0$ corresponding to 100% errors in measuring velocity. We can then compute the average and maximum standard deviation of the estimate U_{est} . As an illustration, the results of these calculations are summarized in Figure 7 for the snapshot 1.

Similarly, the exclusion disk indeed yields smaller standard deviations of the reconstructed as the results shown in Table 4. Here, the results indicate that suitable R also helps to reduce the standard deviation in the prediction. We note that here $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$ is linearly dependent on α . For smaller α , the results can be obtained by simply scaling the numbers in Table 4 and the patterns of the distribution of the deviation is the same as those shown in Figure 7. In the next section, we discuss the case that the measurement error from a sensor is considerably large which means the sensor is malfunctioning.

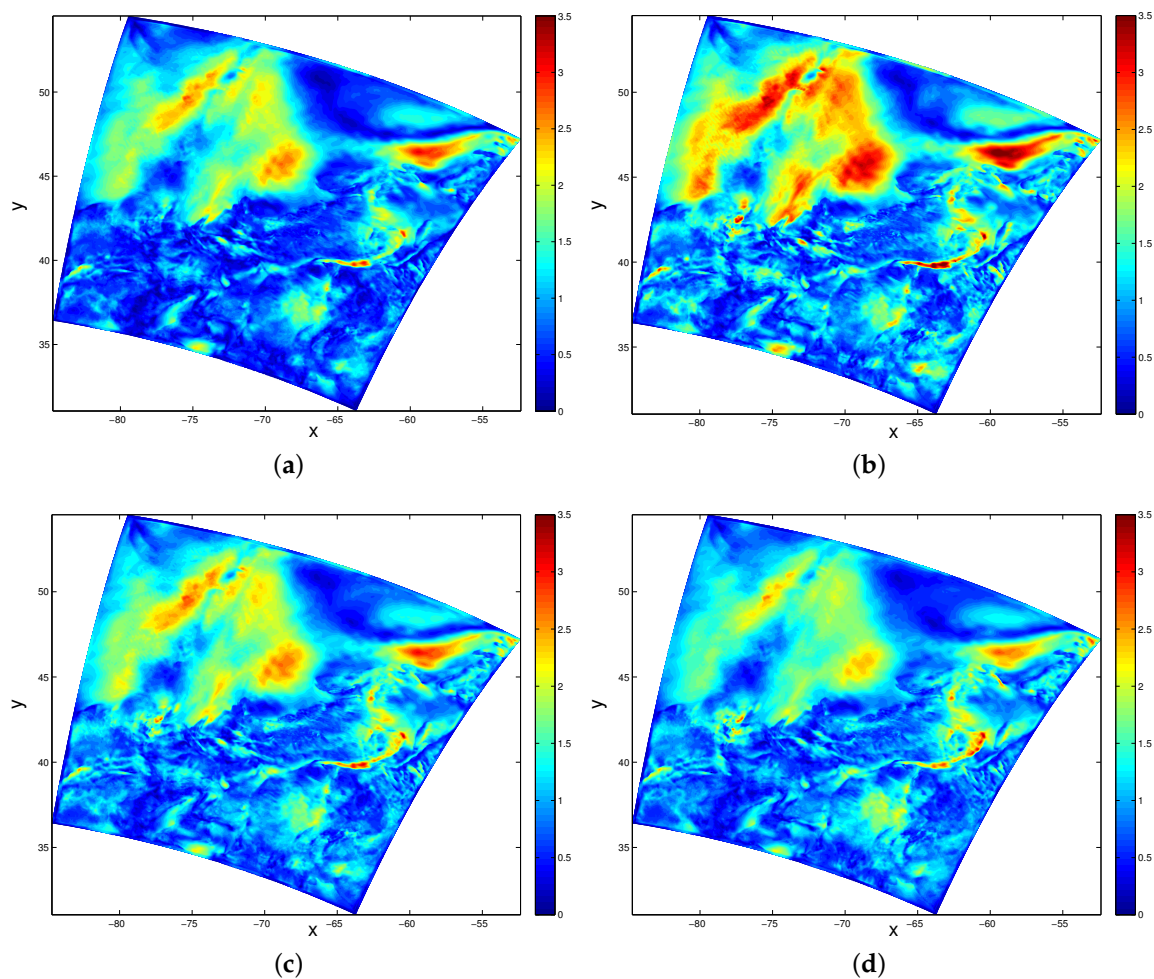


Figure 7. Standard deviation of the reconstructed snapshot 1 based on noisy measurement which depends on the wind field. (a) $R = 0$; (b) $R = 0.5$; (c) $R = 1$; (d) $R = 2$.

Table 4. Average and maximum of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$ over all snapshots given $\sigma_{\xi}(l_i, t) = U(l_i, t)$ for different R .

R Value	$R = 0$	$R = 0.5$	$R = 1$	$R = 2$
maximum of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$	9.147	7.756	5.929	5.820
average of $\sigma_{U_{\text{est}}}(\mathbf{x}, t)$	1.694	1.296	1.294	1.098

4. Detecting the Malfunctioning Sensor

In this section, we assume that one sensor is not functioning correctly, and all the other sensors provide *accurate* data. Specifically, we assume that the observed gappy field for each snapshot is $\tilde{u}(t) = u(t) + \delta_{I^*} u(t)$, where $\delta_{I^*} u(t)$ has only one non-zero entry corresponding to the contaminated sensor at location I^* . The coefficients $\tilde{b}$ for the POD modes minimize the gappy norm $\|\tilde{u} - \tilde{\hat{u}}\|_m$, where $\tilde{\hat{u}}$ is the reconstruction from the Gappy-POD. For each snapshot, given sufficient number of sensors as well as appropriate exclusive disk radius, we have that at each sensor location I_j , $|u(I_j) - \hat{u}(I_j)|^2 \sim \mathcal{O}\left(\frac{1}{N_g} \sum_{k=K+1}^N a_k^2\right)$, where N_g is the number of grid on the computational domain. Hence, if $|\delta_{I^*} u|^2$ is much larger than $\mathcal{O}\left(\frac{1}{N_g} \sum_{k=K+1}^N a_k^2\right)$, we can expect big discrepancy between u and $\hat{u}$ at I^* . We thus propose the following heuristic cross-validation method to detect the malfunctioning sensor :

1. For each sensor location I_i , we use the data from $I_1, I_2, \dots, I_{i-1}, I_{i+1}, \dots, I_{N_s}$ to reconstruct the field and denote the reconstructed value at this point as $\hat{U}(I_i, t)$. Here, we take $N_s = 24$ as we consider only the first 24 snapshot from our data.
2. Compute the difference between the reconstructed value and the observed value at each I_i : $\varepsilon(I_i, t_j) = |\hat{U}(I_i, t_j) - U(I_i, t_j)|$, $j = 1, 2, \dots, N_s$.
3. Compare the sum of $\varepsilon(I_i, t)$ over all snapshots at each I_i . When $\varepsilon_i = \sqrt{\sum_{j=1}^{N_s} \varepsilon^2(I_i, t_j)}$ is large, we claim that the i -th sensor is not functioning well.

Remark 3. The sensors identified by the aforementioned method are candidates for erroneous sensors. Even if only one of the sensors is not working, we can select more than one sensors as candidates for further check. In addition, this method can be applied to the case when more than one sensors are not working.

As a demonstration, we consider a case when $N_s = 24$ sensors are located based on the extrema of the first six POD modes with the exclusion disk of radius $R = 2.0$. To test our cross-validation method, we perturb the measurement from the third sensor with noise and keep the measurements from the rest of sensors unaltered. Moreover, we consider a scenario that this third sensor is working properly from $t = 1$ until $t = 5$, then for the remaining time, the error of this sensor is (1) 25% and (2) 50% for our perturbation. We intend to examine whether our method is capable of identifying the malfunctioning sensor. Figure 8a demonstrates that when the error of the sensor is 25%, ε_3 is apparently much larger than other ε_i , which implies that the third sensor may be malfunctioning (after certain snapshot). By examining $\varepsilon(I_3, t)$ for each t as in Figure 8b, we are also able to identify the time when this sensor starts to malfunction. Figure 9 demonstrates similar results when the error of the sensor is 50%. Comparing Figures 8 and 9, we observe that when the error becomes larger, the anomaly of this third sensor becomes easier to detect with our approach. Notice that more numerical results (not shown here) imply that the cross-validation method works when any of the sensor is malfunctioning. We select the third sensor to be anomaly only for demonstration purpose.

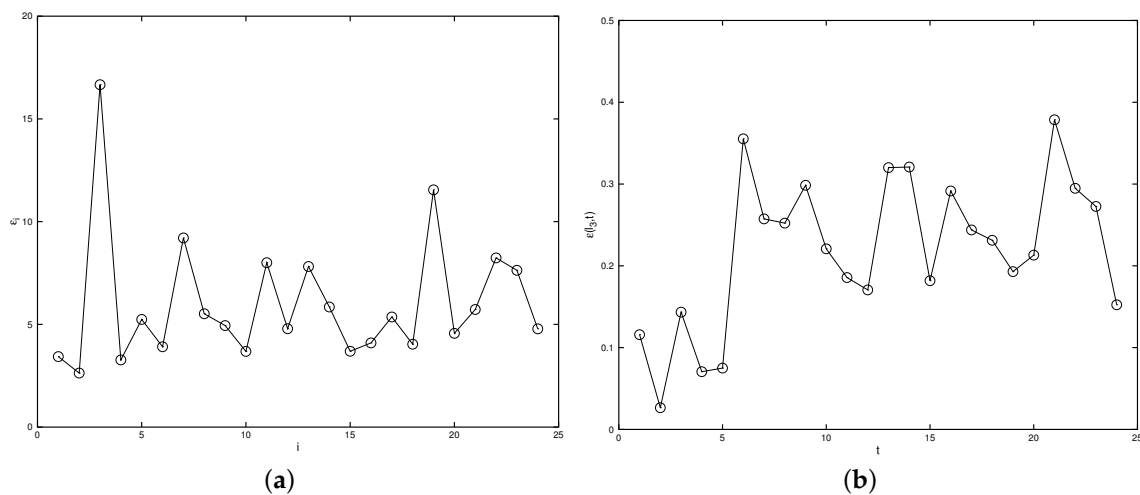


Figure 8. Detecting the malfunctioning sensor by examining $\varepsilon(I_i, t)$ when the error of the sensor is 25%.
(a) ε_i for all i ; (b) $\varepsilon(I_3, t)$ for all t .

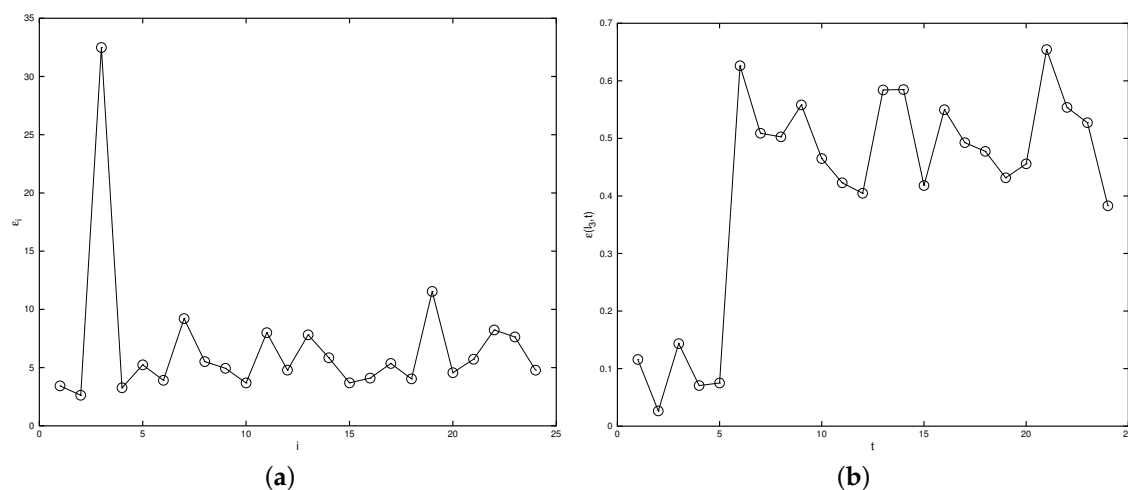


Figure 9. Detecting the malfunctioning sensor by examining $\varepsilon(I_i, t)$ when the error of the sensor is 50%.
(a) ε_i for all i ; (b) $\varepsilon(I_3, t)$ for all t .

5. Conclusions

In this work, we extended the exclusion cylinder strategy to sensor placements for reconstructing two-dimensional wind fields. Our numerical tests suggested that a proper size of area exclusion could improve the quality of the field reconstruction from measurements. We addressed the issue of uncertainty in the measurements by adding Gaussian noise in the 24 hourly spatial wind field simulation data sets in Maine Bay simulated from the WRF model that provided the testbed for our work. In particular, we considered two cases: the uncertainty in the measurements is constant everywhere or is varying in space and time. We checked the standard deviation of the reconstructed field and found proper exclusion disk sizes can reduce the errors of the reconstruction field. With cross-validation techniques, we located a malfunctioning sensor that provides large measurement errors. The numerical results in this paper suggest that the exclusion disk strategy within the Gappy-POD framework can be applied to optimize sensor placements in two-dimensional problems. As future work, we expect to work towards our ultimate goal to develop real-time adaptive sampling

for wind forecasting, which helps to find optimal locations for wind turbines with measurements from functioning sensors.

Acknowledgments: This work is supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program as part of the Multifaceted Mathematics for Complex Energy Systems (M²ACS) project and part of the Collaboratory on Mathematics for Mesoscopic Modeling of Materials project. This work is also partially supported by National Science Foundation (NSF) Grant DMS-1555072. Computations were performed using the National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory. The Pacific Northwest National Laboratory (PNNL) is operated by Battelle for the US Department of Energy under Contract DE-AC05-76RL01830. The first author is supported by a start-up founding from Worcester Polytechnic Institute (WPI).

Author Contributions: Z. Zhang quantified the uncertainty in the measurement and detecting the malfunctioning sensor, Y. Xiu conducted the gappy POD analysis and designed the sensor placement strategy, G. Lin conducted the Weather Research and Forecasting Model simulations and extracted the POD modes from the simulation datasets.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Venturi, D. On Proper Orthogonal Decomposition of Randomly Perturbed Fields with Applications to Flow past a Cylinder and Natural Convection over a Horizontal Plate. *J. Fluid Mech.* **2006**, *559*, 215–254.
2. Rempfer, D. Low-dimensional modeling and numerical simulation of transition in simple shear flow. *Ann. Rev. Fluid Mech.* **2003**, *35*, 229–265.
3. Bekooz, G.; Holmes, P.; Lumley, J. The proper orthogonal decomposition in the analysis of turbulent flows. *Ann. Rev. Fluid Mech.* **1993**, *25*, 539–575.
4. Aubry, N.; Guyonnet, R.; Stone, E. Spatio-temporal analysis of complex signals: Theory and applications. *J. Stat. Phys.* **1991**, *64*, 683–739.
5. Sirovich, L. Turbulence and the dynamics of coherent structures, Parts I, II and III. *Quart. Appl. Math.* **1987**, *XLV*, 561–590.
6. Zhang, Y.; Bellingham, J. An efficient method of selecting ocean observing locations for capturing the leading modes and reconstructing the full field. *J. Geophys. Res.* **2008**, *113*, C04005.
7. Venturi, D.; Karniadakis, G. Gappy Data and Reconstruction Procedures for Flow Past a Cylinder. *J. Fluid Mech.* **2004**, *519*, 315–336.
8. Alvera-Azcárate, A.; Barth, A.; Rixen, M.; Beckers, J.M. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. *Ocean Modelling* **2005**, *9*, 325–346.
9. Beckers, J.; Rixen, M. EOF Calculations and Data Filling from Incomplete Oceanographic datasets. *J. Atmos. Ocean Tech.* **2003**, *20*, 1839–1856.
10. D’Andrea, F.; Vautard, R. Extratropical low-frequency variability as a low dimensional problem. Part I: A simplified model. *Quart. J. Roy. Meteor. Soc.* **2001**, *127*, 1357–1375.
11. Hendricks, J.; Leben, R.; Born, G.; Kobalinsky, C. Empirical orthogonal function analysis of global TOPEX/POSEIDON altimeter data and implications for detection of global sea rise. *J. Geophys. Res.* **1996**, *101*, 14131–14145.
12. Everson, R.; Cornillon, P.; Sirovich, L.; Webber, A. An empirical eigenfunction analysis of sea surface temperatures in the North Atlantic. *J. Phys. Ocean.* **1995**, *27*, 468–479.
13. Wilkin, J.; Zhang, W. Modes of mesoscale sea surface height and temperature variability in the East Australian Current. *J. Geophys. Res.* **2006**, *112*, C01013.
14. Pedder, M.; Gomis, D. Application of EOF Analysis to the spatial estimation of circulation features in the ocean sampled by high-resolution CTD samplings. *J. Atmos. Ocean Tech.* **1998**, *15*, 959–978.
15. Houseago-Stokes, R. Using optimal interpolation and EOF analysis on North Atlantic satellite data. *International WOCE Newsletter* **2000**, *28*, 26–28.
16. Preisendorfer, W.; Mobley, C.D. *Principal component analysis in meteorology and oceanography*; Elsevier: Amsterdam, The Netherlands, 1988.
17. Dickey, T. Emerging ocean observations for interdisciplinary data assimilation systems. *J. Mar. Syst.* **2003**, *40–41*, 5–48.

18. Yildirim, B.; Chrysostomidis, C.; Karniadakis, G.E. Efficient sensor placement for ocean measurement using low-dimensional concepts. *Ocean Modelling* **2009**, *27*, 160–173.
19. Yang, X.; Venturi, D.; Chen, C.; Chrysostomidis, C.; Karniadakis, G.E. EOF-based constrained sensor placement and field reconstruction from noisy ocean measurements: Application to Nantucket Sound. *J. Geophys. Res.* **2010**, *115*, doi:10.1029/2010jc006148.
20. Xue, P.; Chen, C.; Beardsley, R.C.; Limeburner, R. Observing system simulation experiments with ensemble Kalman filters in Nantucket Sound, Massachusetts. *J. Geophys. Res. Oceans* **2011**, *116*, doi:10.1029/2010JC006428.
21. Li, W.; Sun, S.; Jia, Y.; Du, J. Robust unscented Kalman filter with adaptation of process and measurement noise covariances. *Digit. Signal Process.* **2016**, *48*, 93–103.
22. Barrault, M.; Maday, Y.; Nguyen, N.C.; Patera, A.T. An ‘empirical interpolation’ method: Application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique* **2004**, *339*, 667–672.
23. Chaturantabut, S.; Sorensen, D.C. Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.* **2010**, *32*, 2737–2764.
24. Chinesta, F.; Leygue, A.; Bordeu, F.; Aguado, J.; Cueto, E.; González, D.; Alfaro, I.; Ammar, A.; Huerta, A. PGD-based computational vademecum for efficient design, optimization and control. *Arch. Comput. Methods Eng.* **2013**, *20*, 31–59.
25. Nadal, E.; Chinesta, F.; Díez, P.; Fuenmayor, F.; Denia, F. Real time parameter identification and solution reconstruction from experimental data using the Proper Generalized Decomposition. *Comput. Methods in Appl. Mech. Eng.* **2015**, *296*, 113–128.
26. Everson, R.; Sirovich, L. The Karhunen-Loève procedure for gappy data. *J. Opt. Soc. Am., A* **1995**, *12*, 1657–1664.
27. Willcox, K. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Comput. Fluids* **2006**, *35*, 208–226.
28. Bui-Thanh, T.; Damodaran, M.; Willcox, K.E. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *AIAA J.* **2004**, *42*, 1505–1516.
29. Mokhasi, P.; Rempfer, D. Optimized sensor placement for urban flow measurement. *Phys. Fluids* **2004**, *16*, 1758–1764.
30. Cohen, K.; Siegel, S.; McLaughlin, T. Sensor placement based on proper orthogonal decomposition modeling of a cylinder wake. AIAA Paper 2003-4259. In Proceedings of the 33rd AIAA Fluid Dynamics Conference and Exhibit, Fluid Dynamics and Co-located Conferences, Orlando, FL, USA, 23–26 June 2003.
31. Mathelin, L.; Maître, O.L. Robust control of uncertain cylinder wake flows based on robust reduced order models. *Comput. Fluids* **2009**, *38*, 1168–1182.
32. Venturi, D.; Wan, X.; Karniadakis, G.E. Stochastic low-dimensional modelling of a random laminar wake past a circular cylinder. *J. Fluid Mech.* **2008**, *606*, 339–367.

