

# Optimal placement of mobile sensors for data assimilations

By WEI KANG<sup>1</sup> and LIANG XU<sup>2\*</sup>, <sup>1</sup>*Naval Postgraduate School, Monterey, California, USA;*  
<sup>2</sup>*Naval Research Laboratory, Monterey, California, USA*

(Manuscript received 2 January 2012; in final form 11 September 2012)

## ABSTRACT

We explore the theoretical framework as well as the associated algorithms for the problem of optimally placing mobile observation platforms to maximise the improvement of estimation accuracy. The approach in this study is based on the concept of observability, which is a quantitative measure of the information provided by sensor data and user-knowledge. To find the optimal sensor locations, the observability is maximised using a gradient projection method. The Burgers equation is used to verify this approach. To prove the optimality of the sensor locations, Monte Carlo experimentations are carried out using standard 4D-Var algorithms based on two sets of data, one from equally spaced sensors and the other from the optimal sensor locations. The results show that, relative to equally spaced sensors, the 4D-Var data assimilation achieves significantly improved estimation accuracy if the sensors are placed at the optimal locations. A robustness study is also carried out in which the error covariance matrix is varied by 50% and the sensor noise covariance is varied by 100%. In addition, both Gaussian and uniform probability distributions are used for the sensor noise and initial estimation errors. In all cases, the optimal sensor locations result in significantly improved estimation accuracy.

*Keywords:* sensor placement, observability, optimisation, data assimilation, 4D-Var

## 1. Introduction

One of the important components of the numerical weather prediction (NWP) is atmospheric data assimilation. Atmospheric data assimilation provides the best available state estimate (i.e. the initial conditions) to be used in NWP. It is a procedure to combine all the available prior information, such as background, prediction model, observations, observation operators, and the associated error statistics, to obtain the state estimate using proper mathematical techniques. In the same community of data assimilation, various techniques, such as variational algorithms (e.g. 3D-Var and 4D-Var) and ensemble Kalman filters (e.g. EnKF) as well as hybrid variational/ensemble algorithms, are currently used in atmospheric data assimilation.

It is well known within the data assimilation community that not every observation has the same impact on the NWP forecast. The observation impact depends on the information content of each assimilated observation and the data assimilation system used to assimilate it. It is highly desirable to know the potential impact of each

observation on short term NWP forecasts or some other measures of interests. Virtually all the existing researches in this area are based on the pioneering work of Baker and Daley (2000). They introduced the concept of observation sensitivity (also known as the gradient of atmospheric analysis with respect to the observations). Baker and Daley (2000) showed that observation sensitivity is simply the adjoint of the data assimilation system. They also provided an elegant way to construct the adjoint of an observation-space variational data assimilation system. Langland and Baker (2004) proposed a technique to routinely calculate the observation impact on short term NWP forecasts (24 hr forecast in their case). The technique requires: a) the adjoint of the US Navy's global NWP model (NOGAPS, see Hogan and Rosmond (1991); Rosmond (1997)) to provide the sensitivity of the 24 hr forecast errors with respect to the initial condition; b) the adjoint of the US Navy's 4D-Var data assimilation system [NAVDAS-AS, see Xu et al. (2005)]. Their results provided very useful inside information that was impossible to obtain before.

Because observations play such an important role in the data assimilation and in the subsequent forecast, many additional studies to address various aspects of the observation impact have been conducted since 2000

---

\*Corresponding author.  
email: liang.xu@nrlmry.navy.mil

at various institutions around the world. The following is a list of works related to the variational data assimilation algorithms: Cardinali et al. (2004); Xu et al. (2006); Daescu (2008); Tremolet (2008); Baker and Langland (2009); Cardinali (2009); Daescu and Todling (2009, 2010); Gelaro and Zhu (2009); Gelaro et al. (2010); Cardinali and Prates (2011). Similar studies were also conducted for ensemble data assimilation methods, such as Liu and Kalnay (2008), Liu et al. (2009), and Li et al. (2009). Despite the success in the area of observation impact, this can only provide the observation impact information after the observation is already assimilated and the analysis is used as the initial condition to make the short term forecast. Currently it is not possible to deal with the issue of optimal sensor placement in the planning of the targeting observation during a field experiment.

In this study, we explore the theoretical framework as well as the associated algorithms for the problem of optimally placing mobile observation platforms to maximise the improvement of estimation accuracy. The current adjoint (or ensemble) based observation monitor system can provide the evaluation of the observation impact only after the fact. A comprehensive solution to the planning of optimal sensor placement is still an open problem. In our approach, the optimal sensor locations are calculated around a nominal trajectory or a forecast defined by an initial state  $u_0$ . This is necessary because, for non-linear systems, the observability is trajectory dependent. Using the same sensor locations, the observability can have significant change around different trajectories. For sensors with mobility, the issue of unknown nominal trajectory is resolved using a moving horizon approach. The optimal sensor locations are found based on the current approximate of the system trajectory. After a period of time, an updated estimate of the system trajectory is computed and a forecast is given. Based on that, a new set of optimal sensor locations are found and the sensors are relocated. The process is repeated in a time interval synchronised with the estimation and forecast process.

The theory and algorithm developed in this study are based on the concept of observability, which is a quantitative measure of the information provided by sensor data and user-knowledge. The observability can be numerically computed based on the system's dynamic model. It provides the cost function for the optimisation of sensor locations. In this study, the Burgers equation is used to verify the theory and the approach. In the computation, the system's observability is numerically approximated using an empirical Gramian matrix method (to be explained below). Then the observability is maximised using a gradient projection method to find the optimal sensor locations. To verify the optimality of the method, Monte Carlo numerical experiments are carried out using standard

4D-Var algorithms based on two sets of data, one from equally spaced sensors and the other from the optimal sensor locations. The results show that, relative to equally spaced sensors, the 4D-Var data assimilation achieves significantly improved estimation accuracy if the sensors are placed at the optimal locations. A robustness study is also carried out in which the initial error covariance matrix is varied by 50% and the sensor noise covariance is varied by 100%. To further test the robustness, the numerical experiments are implemented using various sets of data with both Gaussian and uniform probability distributions. In all cases, the optimal sensor locations result in significantly improved estimation accuracy.

The study is organised as follows. In Section 2, the concept of observability is introduced for general dynamical systems. Its approximation using an empirical Gramian matrix is also introduced. In Section 3, the problem of optimal sensor placement is formulated and a gradient projection method is introduced for its computation. In Section 4, the concept and the algorithms are applied to a system defined by the Burgers equation as an illustrative example. The optimal sensor locations are found. In Section 5, a standard 4D-Var data assimilation method is applied to test the optimal sensor locations. In Section 6, the robustness of the optimal sensor locations is tested using several covariance matrices as well as different probability distributions. At the end of this section, we revisit the definition of observability to address some related issues on the norms and metrics used in the concept.

## 2. Observability

The concept of observability has been widely used in control system theory and its applications. In this study, we follow the recent research work by Kang and Xu (2009a, 2009b) on the quantitative measure of observability. Consider a discrete-time dynamical system

$$u(k+1) = f(k, u(k)) \quad (1)$$

where  $u \in \mathbb{R}^{N_s}$  is the state variable of the system,  $k$  is an integer representing the time steps since the initial time, and  $f$  is a function of  $(k, u)$ . If the original model is defined by a continuous-time differential equation, then (1) represents its discretisation where  $u(k)$  is the state of the system at grid points. Let  $\lambda$  be a parameter that defines the sensor configuration. In this study,  $\lambda$  represents the locations of the sensors in space. However, the concept is applicable to other sensor parameters such as path, orientation, and density. The output, denoted by  $y(k)$ , of eq. (1) is the variable measured by sensors. It is a function of the state variable and sensor locations,

$$y(k) = h(u(k), \lambda), y \in \mathbb{R}^{N_y}$$

In the simulations,  $h(u, \lambda)$  is computed using spline if  $\lambda$  is not located at grid points. Other interpolating algorithms are applicable as well. To summarise, we study the observability of the state variable based on the following system model:

$$\begin{aligned} u(k+1) &= f(k, u(k)), & u &\in \mathbb{R}^{N_u} \\ y(k) &= h(u(k), \lambda), & y &\in \mathbb{R}^{N_y} \\ 0 &\leq k \leq N_t \end{aligned} \quad (2)$$

Let  $u(k; u_0)$ ,  $k=0, 1, \dots, N_t$ , be the trajectory of the system with initial state  $u(0; u_0) = u_0$ . The corresponding output is denoted by  $y(k; \lambda; u_0)$ , i.e.

$$y(k, \lambda; u_0) = h(u(k; u_0), \lambda)$$

In the following, we define the observability of  $u(0)$ . The definition can be generalised to the observability of the state at any time,  $k=k_0$ . Consider a nominal trajectory  $u(k; u_0)$ ,  $k=0, 1, \dots, N_t$ . The variation of the measured variable  $y(k)$  under the variation of  $u_0$  is defined as follows:

$$\begin{aligned} &J(u_0, \delta u_0, \lambda) \\ &= \delta u_0^T P_1 \delta u_0 + \|y(\cdot, \lambda; u_0 + \delta u_0) - y(\cdot, \lambda; u_0)\|_Y \end{aligned} \quad (3)$$

where  $P_1$  is a weight matrix that reflects our confidence on the background from the previous estimation process step or the initial guess. If  $P_1$  is large, the background is most likely close to the true value, which implies a good observability. Similar to data assimilation,  $P_1$  can be assigned as the inverse matrix of the background error covariance. In this study, the term including  $P_1$  is not applied so that the study is focused on the second term in eq. (3), which represents the impact of the sensor information. We adopt the following norm for  $\|\cdot\|_Y^2$ .

$$\sum_{k=0}^{N_t} \|y(k, \lambda; u_0 + \delta u_0) - y(k, \lambda; u_0)\|_{P_2}^2$$

where  $\|v\|_{P_2}^2 = v^T P_2 v$ . Because  $y(\cdot)$  represents the output from various types of sensors with different error probability distributions,  $P_2$  in eq. (3) is a weight matrix, for instance, the inverse matrix of the sensor noise covariance. In addition,  $P_2$  can be used to weight the importance of different sensors in a data assimilation process, which is a potential application of the concept to sensor assessment.

The state space of the system is  $\mathbb{R}^{N_u}$ . However, it is not necessarily the space used for the estimate, or the space of estimation. For large scale systems with high dimension  $N_u$ , it is often the case that a low dimensional subspace is able to provide adequate estimation accuracy. For the definition of observability, we introduce the space for estimation,  $W \subseteq \mathbb{R}^{N_u}$ . The space has a norm,  $\|\cdot\|$ , which is the same norm that we use to measure the error of an estimated  $u_0$ . We assume that the estimate is updated at each time step using vectors in  $W$ . The following definition of observability is the discrete-time version of a special case defined in

Kang and Xu (2009a). It agrees in spirit with the approach in Krener and Ide (2009).

**Definition 1.** Let  $\rho > 0$  be a positive number. Then the number  $\varepsilon$  is defined as follows

$$\begin{aligned} \varepsilon^2 &= \min_{\delta u_0} J(u_0, \delta u_0, \lambda) \\ &\text{subject to} \\ &\|\delta u_0\| = \rho, \\ &\delta u_0 \in W \end{aligned} \quad (4)$$

The ratio  $\rho/\varepsilon$  is a measure of observability. It is called an unobservability index.

The value of  $\varepsilon$  represents the smallest variation of  $y$  corresponding to variations of  $u_0$ . A small value of  $\varepsilon$  implies that  $u_0$  is less observable. In the extreme case of  $\varepsilon=0$ , it implies that the sensor cannot detect the variation of  $u_0$  in at least one direction; and  $P_1$  does not contain any information about the accuracy of the initial guess in the same direction. Therefore,  $u_0$  is unobservable. The value of  $\delta u_0$  obtained from solving the optimisation problem [eq. (4)] represents the direction of  $\delta u_0$  in which the sensors and initial guess are least sensitive or accurate. To summarise, a small value of  $\rho/\varepsilon$  implies strong observability. For the case of linear systems with  $w = \mathbb{R}^{N_u}$ , the ratio  $\rho/\varepsilon$  equals the value of unobservability index defined in Krener and Ide (2009). If we consider  $J$  in eq. (3) as the error of sensors and  $\|\delta u_0\|$  as the estimation error, then  $\rho/\varepsilon = 1$  implies that the magnitude of the worst error in the estimation of  $u_0$  is at about the same level as the sensor error. If  $\rho/\varepsilon$  belongs to the interval  $[1, 10]$  then the worst estimation error is larger than the sensor error, but in the same order of magnitude. Depending on the error tolerance required by data assimilation, for specific systems one can define the ranges of the value of  $\rho/\varepsilon$  in which the system is strongly observable, reasonably observable, weakly observable, and unobservable.

Interested readers are referred to Kang and Xu (2009b) for illustrative examples of observability. The following is an artificial and simple example of an unobservable system. In this case, the unobservability index is  $\infty$ . Suppose

$$\begin{aligned} u_1(k+1) &= u_1(k) + u_2(k) \\ u_2(k+1) &= u_2(k) + u_3(k) \\ u_3(k+1) &= u_3(k) + 1 \end{aligned} \quad (5)$$

Suppose the sensors are able to measure the value of  $u_2$ , i.e.

$$Y(k) = u_2(k)$$

Given any initial condition  $u_0 = [u_{01} \ u_{02} \ u_{03}]^T$ , we can generate two trajectories using  $u(0) = u_0$  and  $u(0) = u_0 + [0.1 \ 0 \ 0]^T$ , i.e.  $\rho = \|\delta u_0\| = 0.1$ . It is obvious that the two different trajectories of  $u(k)$  result in the same value of  $y(k)$ . In other words, the sensor completely fails to catch the change of value in the initial condition. In this case,  $\varepsilon=0$  because the cost in eq. (3) equals zero, where we assume

$P_1=0$  (no a priori statistical information about the initial condition). Obviously  $u_1(0)$  is not observable. The unobservability index,  $\rho/\varepsilon$ , equals infinity because  $\varepsilon=0$ .

The system becomes observable if we can measure  $u_1(k)$ , i.e.

$$y(k) = u_1(k)$$

The system [eq. (5)] with this output can be written into the following matrix form

$$\begin{aligned} u(k+1) &= Au(k) + B \\ y(k) &= Cu(k) \end{aligned} \quad (6)$$

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$C = [1 \quad 0 \quad 0]$$

If  $u(k)$  is a trajectory of eq. (6) with an initial state  $u_0$ , then the trajectory with an initial variation  $\delta u_0$  is

$$u(k) + A^k \delta u_0$$

Therefore

$$y(k; u_0 + \delta u_0) - y(k; u_0) = CA^k \delta u_0$$

If we use the Euclidean norm in eq. (3), then

$$\begin{aligned} J^2 &= \sum_{k=0}^N \delta u_0^T (A^k)^T C^T CA^k \delta u_0 \\ &= \delta u_0^T G \delta u_0 \end{aligned}$$

where  $N+1$  is the total number of the sampling data used in the estimation process, and

$$G = \sum_{k=0}^N (A^k)^T C^T CA^k$$

Therefore,  $\varepsilon$  in eq. (4) satisfies

$$\varepsilon^2 = \sigma^2 \|\delta u_0\|_2^2$$

where  $\sigma^2$  is the smallest eigenvalue of  $G$ . Suppose  $\|\delta u_0\|_2 = \rho$ , then

$$\rho/\varepsilon = 1/\sigma$$

Following this example, we can approximate the unobservability index for non-linear systems with a small value of  $\rho$ . The matrix  $G$  can be numerically approximated by a method of empirical Gramian matrix, which is a generalisation of the algorithm in Krener and Ide (2009). Consider the mapping

$$\delta u_0 \rightarrow (\delta u_0, y(\cdot, \lambda; u_0 + \delta u_0) - y(\cdot, \lambda; u_0)) \quad (7)$$

From eq. (3), the cost function  $J(u_0, \delta u_0, \lambda)$  is, in fact, the norm of the image of this mapping. The minimisation in eq. (4) is to find the image of the mapping [eq. (7)] with smallest norm subject to  $\|\delta u_0\| = \rho$ . Let  $v_1, v_2, \dots, v_s$  be

an orthonormal basis of the space  $W$ . Then the variation of  $\delta u_0$  on the sphere  $\|\delta u_0\| = \rho$  can be represented by

$$\delta u_0 = \sum \alpha_i v_i$$

where

$$\sum \alpha_i^2 = \rho^2$$

For each direction  $v_i$ , the variation of the output can be estimated empirically by

$$\Delta_i(k) = \frac{1}{2\rho} (y(k, \lambda; u_0 + \rho v_i) - y(k, \lambda; u_0 - \rho v_i)), \quad (8)$$

for  $k=0, \dots, N_i$ . The mapping [eq. (7)] can be locally approximated by

$$\delta u_0 = \sum \alpha_i v_i \rightarrow (\delta u_0, \sum \alpha_i \Delta_i(\cdot)) \quad (9)$$

Therefore, the minimum value in eq. (4) approximately equals the smallest norm of the images under the mapping [eq. (9)] subject to  $\sum \alpha_i^2 = \rho^2$ . More specifically,

$$J(u_0, \delta u_0, \lambda) \approx \alpha^T G \alpha, \quad \alpha = [\alpha_1 \quad \dots \quad \alpha_s]^T \quad (10)$$

where

$$\begin{aligned} G &= (G_{ij})_{i,j=1}^s \\ G_{ij} &= \langle v_i, v_j \rangle_{P_1} + \langle \Delta_i, \Delta_j \rangle_{P_2} \\ &= v_i^T P_1 v_j + \sum_{k=0}^{N_i} \Delta_i(k)^T P_2 \Delta_j(k) \end{aligned}$$

Let  $\sigma^2$  be the smallest eigenvalue of  $G$ , then the smallest norm of the images under eq. (9) subject to  $\sum \alpha_i^2 = \rho^2$  is  $\sigma\rho$ . Therefore, the unobservability index  $\rho/\varepsilon$  is approximated by  $1/\sigma$ .

In linear algebra,  $G$  is called a Gramian matrix. Because we are trying to compute the observability before sensors collecting data and because of the random sensor error, there are endless possible estimates of a state variable. The Gramian matrix  $G$  approximates the mapping from the possible estimation error to the variation of the output, or the possible sensor error.

What eq. (4) measures is, in fact, the observability of the initial state  $u_0$ . However, an accurate initial state does not automatically guarantee good accuracy along the entire trajectory. One may want to measure the observability using  $u(N_i)$ , the final state, or other time point  $u(k)$ , to define the observability. Because a trajectory is uniquely determined by any given  $u(k)$  for a fixed time  $k$ , it is straightforward to generalise Definition 1 to the observability of any  $u(k)$ . The selection of  $k$  depends on the system or the accuracy requirement of specific applications. To guarantee an overall accuracy, it is also possible to use different norms, such as  $L^2$ -norm. This issue is addressed in Section 6.

### 3. Optimal sensor placement

The concept of observability provides a quantitative measure of the quality of sensor information for the purpose of state estimation. More specifically, the value of  $\varepsilon$  defined in eq. (4) depends on  $\lambda$ , the sensor locations. We treat  $\varepsilon$  as a function of  $\lambda$ ,  $\varepsilon(\lambda)$ , which is used as the performance measure of sensor locations. The best sensor locations are those that maximise the value of this performance measure, i.e.

$$\begin{aligned} & \max_{\lambda} \varepsilon(\lambda) \\ & \text{subject to} \\ & \lambda_{\min} \leq \lambda \leq \lambda_{\max} \end{aligned} \quad (11)$$

From Section 2, the value of  $\varepsilon(\lambda)$  can be computed using the method of empirical Gramian matrix. Therefore, eq. (11) is, in fact, a problem of maximisation with both linear and non-linear constraints. This problem is numerically challenging. Further research is required to find efficient computational methods. For the purpose of proving the concepts of observability and sensor placement, in the following we adopt a projection gradient method to numerically solve eq. (11).

The gradient with respect to  $\lambda$  is defined by

$$\nabla_{\lambda} = \frac{\partial \varepsilon}{\partial \lambda}$$

Suppose the location of the  $i$ th sensor is bounded by  $\lambda_{\min,i} \leq \lambda_i \leq \lambda_{\max,i}$ . To find the searching direction on the boundary of the region, the following projected gradient is used.

$$\overline{\nabla}_{\lambda,i} = \begin{cases} 0, & \text{if } \lambda_i = \lambda_{\min,i} \text{ and } \nabla_{\lambda,i} < 0 \\ 0, & \text{if } \lambda_i = \lambda_{\max,i} \text{ and } \nabla_{\lambda,i} > 0 \\ \nabla_{\lambda,i}, & \text{otherwise} \end{cases}$$

where  $\nabla_{\lambda,i}$  is the  $i$ th component of  $\nabla_{\lambda}$ . The projected gradient direction guarantees that the searching direction points toward the inside of the boxed region. Thus eq. (11) can be treated as a problem of unconstrained maximisation. Given a search direction  $\overline{\nabla}$ , a process of line search is then carried out using the Armijo algorithm model. Details about this line search algorithm are referred to Polak (1997).

### 4. An example of Burgers' equation

To verify the concept, the theory is applied to the following Burgers' equation

$$\begin{aligned} \frac{\partial U(x,t)}{\partial t} + U(x,t) \frac{\partial U(x,t)}{\partial x} &= \kappa \frac{\partial^2 U(x,t)}{\partial x^2} \\ U(x,0) &= U_0(x) \\ U(0,t) &= f_1(t) \\ U(L,t) &= f_2(t) \end{aligned} \quad (12)$$

Suppose we can measure the value of  $U(x,t)$  at  $N_y$  fixed locations in the time interval  $[0,T]$ . The question is: what sensor locations maximise the observability around the nominal trajectory  $U(x,t)$ ?

For the purpose of computation, we discretise the problem at a set of uniformly spaced nodes,

$$x_0 = 0 < x_1 < \dots < x_N = L,$$

where

$$\Delta x = x_{i+1} - x_i = L/N.$$

Sensors are located in the interval  $[0,L]$ , not necessarily at the nodes (Fig. 1). Define  $u_i(t) = U(x_i,t)$  for  $i = 1, 2, \dots, N-1$ . System [eq. (12)] is discretised using the central difference method

$$\begin{aligned} \dot{u}_1 &= -u_1 \frac{u_2 - u_1(t)}{2\Delta x} + \kappa \frac{u_2 + f_1(t) - 2u_1}{\Delta x^2} \\ \dot{u}_2 &= -u_2 \frac{u_3 - u_1}{2\Delta x} + \kappa \frac{u_3 + u_1 - 2u_2}{\Delta x^2} \\ &\vdots \\ \dot{u}_{N-1} &= -u_{N-1} \frac{f_2(t) - u_{N-2}}{2\Delta x} + \kappa \frac{f_2(t) + u_{N-2} - 2u_{N-1}}{\Delta x^2} \end{aligned} \quad (13)$$

Using the framework in Section 2, the discretised version of eq. (13) represents the dynamical system [eq. (2)]. In all computations that follow, we apply a Runge-Kutta algorithm to eq. (13). The observability is computed around a nominal trajectory defined by the initial value

$$u_0 = [U_0(x_1) \quad U_0(x_2) \quad \dots \quad U_0(x_{N-1})]^T \quad (14)$$

In the following, we use eq. (13) as the true model. The locations of the  $N_y$  sensors are represented by the parameter

$$\lambda = [\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_{N_y}]$$

where  $\lambda_i$  is the  $x$ -coordinate of the sensor in space. A solution of the system in the  $(t,x,u)$ -space with the points of sensor sampling is shown in Fig. 2.

We assume that sensors can be located anywhere in the interval  $[0,L]$ , not necessarily at the nodes. Because of that, the value of the true solution at the sensor locations is not directly available from the discretised model. For the purpose of computation, the value of sensor data is generated by the following interpolation

$$\begin{aligned} y &= h(u, \lambda) \\ &= \text{In}([x_1 \dots x_{N-1}], u, \lambda) \end{aligned} \quad (15)$$

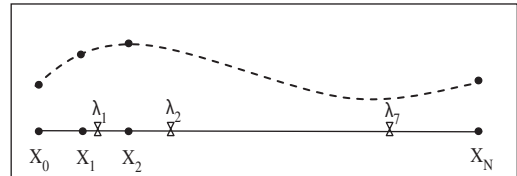


Fig. 1. Illustrative graph of nodes in discretisation and sensor locations.

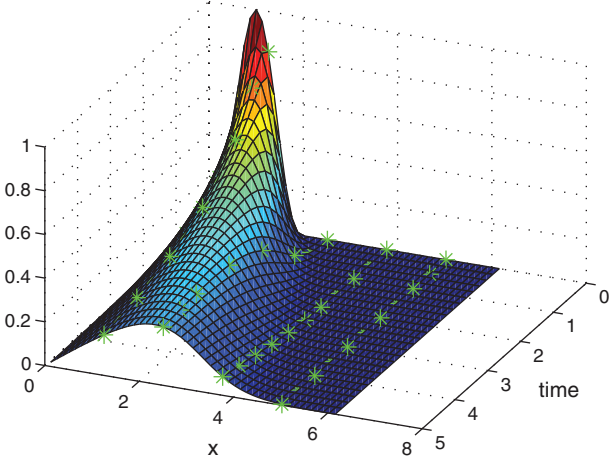


Fig. 2. A trajectory with sensor locations.

where  $\text{In}(\alpha, \beta, \gamma)$  represents the interpolation of  $(\alpha, \beta)$  evaluated at  $\gamma$ . Because  $\lambda$  is a vector,  $h(u, \lambda) \in \mathbb{R}^{N_y}$ . In this study, the interpolations are computed using cubic splines. In the framework introduced in Section 2, the output,  $y$ , of the system in eq. (2) is defined by eq. (15).

Suppose each sensor collects  $N_t$  readings at the frequency  $1/\Delta t$  in the time interval  $[0, T]$ , i.e.  $\Delta t = T/N_t$ . The cost function defined in eq. (3) with constant weights has the following expression

$$\begin{aligned} J(\delta u_0, \lambda) &= \sum_{k=0}^{N_t} \|h(u(t_k; u_0), \lambda) - h(u(t_k; u_0 + \delta u_0), \lambda)\|^2 \\ &= \sum_{k=0}^{N_t} \|\text{In}(X, u(t_k; u_0), \lambda) - \text{In}(X, u(t_k; u_0 + \delta u_0), \lambda)\|^2 \end{aligned} \quad (16)$$

where  $X = [x_1 \cdots x_{N-1}]$ ,  $t_k = t_{k-1} + \Delta t$ . In this example, the cost does not include the norm of the initial error, i.e. we focus on the observability with respect to the sensor data only. The minimisation in eq. (4) has the following form

$$\left\{ \begin{array}{l} \varepsilon^2 = J(\delta u_0, \lambda) \\ \text{subject to} \\ \|\delta u_0\| = \rho_o \\ \delta u_0 \in W \\ \dot{u} = f(u), f(u) \text{ is defined in (13)} \end{array} \right.$$

The observability is measured by the ratio  $\rho/\varepsilon$ . The smaller is this ratio, the stronger is the observability.

The computations are based on a set of specific parameter values. The number of sensors is seven, i.e.  $N_y = 7$ . The nodes for discretisation divide the space,  $[0, 2\pi]$ , into  $N = 50$  equal length subintervals. However, the dimension of the space for estimation,  $W$ , is much smaller. In fact,

we assume that the initial state  $U(x, 0)$  can be accurately approximated using finite terms of Fourier series,

$$U(x, 0) \approx \sum_{k=1}^6 \alpha_k \cos(kx) + \beta_k \sin(kx) \quad (17)$$

The space for estimation, i.e.  $W$  in Definition 1, is defined to be all functions in the form of eq. (17). The observability is defined around a nominal trajectory. In this example, the nominal trajectory is defined by the initial state,  $U_0(x)$ , which is given below. The plot of the nominal trajectory is shown in Fig. 2. We assume that  $U(x, t)$  can be approximated by Fourier series. Therefore, the space for estimation is

$$W = \text{span} \left\{ \alpha_0 + \sum_{k=1}^{N_F} \alpha_k \cos(kX) + \beta_k \sin(kX) \mid \alpha_k, \beta_k \in \mathbb{R} \right\}$$

where  $X = [x_1 \cdots x_{N-1}]^T$ . The initial value of the truth is

$$U_0(x) = \begin{cases} x^3(2-x)^3, & x \leq 2 \\ 0, & x > 2 \end{cases}$$

Details on the parameters used in this example are listed as follows.

$$\begin{array}{ll} N_y = 7 & \text{(seven sensors)} \\ N_F = 6 & \\ \kappa = 0.14 & \\ L = 2\pi & \text{(length of } x\text{-interval)} \\ N = 50 & \text{(} N - 1 \text{ is the dimension of (13))} \\ T = 5 & \text{(final time)} \\ N_t = 20 & \\ \Delta t = T/N_t & \text{(time step size of sensors)} \\ f_1(t) = 0 & \text{(boundary condition)} \\ f_2(t) = 0 & \text{(boundary condition)} \\ \rho_o = 0.01 & \text{(radius of the variation of } u_0) \end{array}$$

Following the formulation and algorithm in Section 3, the optimal sensor locations are computed under the constraints  $2\pi/N \leq \lambda_i \leq 2(N-1)\pi/N$ ,

$$\lambda = [0.7504 \ 1.0628 \ 1.7664 \ 2.0812 \ 3.9754 \ 4.8250 \ 5.4302] \quad (18)$$

In this case, the unobservability index is

$$\rho/\varepsilon = 1.75$$

This number implies that, in the worst case, the error of the  $u(0)$  estimate is about 1.75 times the sensor error as measured by the norm  $J(\delta u_0, \lambda)$ . In other words, a reasonable estimate of  $u(0)$  should have an error within 1.75 times the sensor error. Because of different scales used for the state space and sensor errors, the value of unobservability index must be combined with the expected, or standard deviation, of the sensor error to assess the accuracy of state estimate.

If the sensor error is at the same scale as the desired bound of estimation error, then  $\rho/\varepsilon=1.75$  implies that  $u(0)$  is observable. On the other hand, independent of the scales of the variables, the unobservability index is directly comparable to other sensor locations. For the purpose of comparison, let us consider the equally spaced sensor locations,

$$[\lambda_1 \ \lambda_2 \ \dots \ \lambda_7] = \left[ \frac{2\pi}{8} \ 2\frac{2\pi}{8} \ \dots \ 7\frac{2\pi}{8} \right] \quad (19)$$

In this case, the unobservability index is

$$\rho/\varepsilon = 12.92$$

It is larger than that from optimal sensor locations by more than seven times. We can conclude that the state variable is weakly observable using equally spaced sensors. Its data is much less valuable than those from optimal sensor locations. Therefore we expect significantly larger error in state estimation if sensors are equally spaced, no matter what estimation method is used.

The equally spaced and the optimal sensor locations are shown in Fig. 3. The optimal sensor locations agree with our intuitive sense. It is not the full values of the model state trajectory  $U(x,t)$  that are important for the observability, but rather the dynamics of the trajectory. As shown in Fig. 2,  $U(x,t)$  is much more dynamically active with relatively high rate of change on the left side in the space. On the right side,  $U(x,t)$  is relatively calm. Optimising the observability automatically takes this phenomenon into consideration. As a result, the optimal solution places four sensors on the left side; and leaves three sensors on the right side. Because the observability is relative to nominal trajectories based on forecast, it is desirable that at least some sensors have mobility so that the optimal sensor

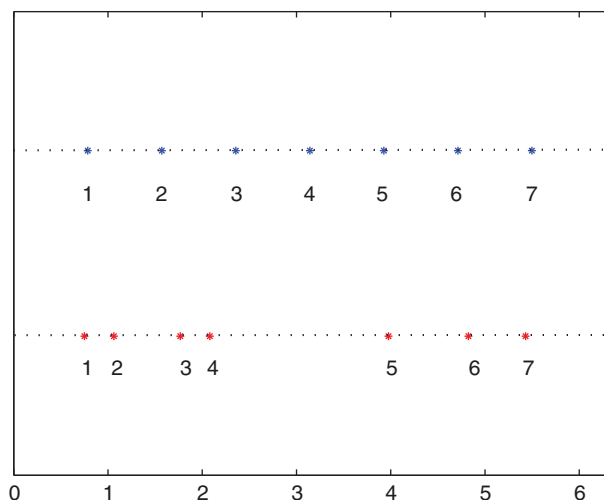


Fig. 3. Sensor locations: equally spaced (blue) and optimal locations (red).

locations can be updated in real-time and the sensor system is adaptive to the dynamic reality.

## 5. Optimal sensor locations for data assimilation

High observability of a trajectory implies that a reasonable estimation method should be able to recover the trajectory with small error. More specifically, if the estimation method is based on the minimisation of the cost function  $J$  in eq. (3), then the optimal sensor locations guarantee to minimise the worst error. Data assimilation is based on the minimisation of a cost function. We expect that the concept of observability is applicable to data assimilations. More specifically, we want to verify the following belief: improved observability results in smaller estimation error. For this purpose, data assimilation results using optimal sensor locations are compared to the results from equally spaced sensors. The comparison is based on a set of Monte Carlo experiments using 200 data sets consisting of randomly generated initial background errors and sensor noise.

For each set of data, a standard algorithm of 4D-Var is applied to estimate the trajectory. For the reasons of space, details on the 4D-Var algorithm used in this section are referred to Xu and Daley (2000) and Xu et al. (2007). The tangent linear model and the adjoint model are developed using the algorithm from Rosmond (1997). The linear equations in the data assimilation are solved using a standard conjugate gradient algorithm in Golub and Van Loan (1996).

The data sets for the Monte Carlo experiments are generated randomly. Two hundred sets of background trajectories are generated around the true trajectory with initial state  $u_0$  in eq. (14). The background,  $u^b(t)$ , represents an estimate from the previous step or an initial guess of the true trajectory. The variation of the initial state of backgrounds,  $u^b(0) - u_0$ , is a random variable with a Gaussian distribution of zero mean. The average of  $\|u_0^b - u_0\|$  is 0.3638, which is about 11.14% relative to  $\|u_0\|$ . The data assimilation uses backgrounds that are projected to the space for estimation,  $W$ , defined in eq. (4).

In a 4D-Var data assimilation, an estimate  $u^a(t)$  is obtained using the system model as well as sensor information, i.e. the output  $y(t_i)$ , which is corrupted by noise. For each background, we generate a set of sensor information with noise

$$y(t_i) = y^{\text{true}}(t_i) + Rv_i$$

$$v_i \in \mathbb{R}^{N_y} - \text{standard white Gaussian noise} \quad (20)$$

$$R = 0.001$$

where  $y^{\text{true}}(t_i)$  is the value of the true state at the sensor location. The standard deviation of the sensor noise,  $R$ , is about 0.7 ~ 1% of the average value of  $u(x,t)$ . Later in the robustness study, it is increased to about 2%. We choose

a relatively small noise because this study is specifically focused on the impact on data assimilation by the single factor of sensor locations.

The sensors are not always placed at the grid points of the discretised model [eq. (13)]. Therefore, in the simulations  $y^{\text{true}}(t_i)$  is not directly available from the model. In the computation of observability, cubic spline is used to approximate its value, which is defined in eq. (15). However, for 4D-Var estimation we adopt a linear interpolation method to approximate  $y^{\text{true}}(t_i)$  so that the matrix  $H$  in the algorithm is easily available. Numerical experiments (results are not shown) confirm that the error caused by the linear interpolation is negligible.

In the Monte Carlo experiments, 200 sets of background are generated as described above,

$$\{u_k^b(t) | k = 1, 2, \dots, 200\} \quad (21)$$

The notation in this section is different from previous sections. The sub-index,  $k$ , in  $u_k^b(t)$  is not the time step as in previous sections. It represents the index in the Monte Carlo experiments. The time step is represented by  $t_i$ . For each background, two sets of sensor data with noise are generated, one based on equally spaced sensors and the other on optimal sensor locations. The sensor data is available only at the sampling point in time and space, i.e. a sequence of the following form

$$\begin{bmatrix} y_1(t_0) & y_1(t_1) & \dots & y_1(t_{N_t}) \\ y_2(t_0) & y_2(t_1) & \dots & y_2(t_{N_t}) \\ \dots & \dots & \dots & \dots \\ y_{N_y}(t_0) & y_{N_y}(t_1) & \dots & y_{N_y}(t_{N_t}) \end{bmatrix} \quad (22)$$

where  $N_t = 20$  and  $N_y = 7$ . For each combination of background and sensor locations, the analysis,  $u^a(t)$ , is found using 4D-Var data assimilation. The analysis is an estimate of the truth. Two norms are used in the evaluation of estimation error from different sensor locations. Let  $u^{\text{truth}}(t)$  be the true trajectory of the system, i.e. the truth. For an analysis,  $u^a(t)$ , the overall error of  $u^a(t)$  is measured using the following norm

$$\|u^a - u^{\text{truth}}\|^2 = \int_0^T \|u^a(t) - u^{\text{truth}}(t)\|^2 dt \quad (23)$$

where  $\|u^a(t) - u^{\text{truth}}(t)\|$  is the norm in  $\mathbb{R}^{N-1}$ . The error of the initial state  $u^a(0)$  is defined by

$$\|u^a(0) - u_0\| \quad (24)$$

Using the backgrounds in eq. (21) and the 4D-Var estimation process, 200 estimate of the truth is numerically generated. The accuracy of the data assimilation is

characterised by the root-mean-square error (RMSE) based on the norms eq. (23) and eq. (24),

$$\sqrt{\frac{\sum_{k=1}^N \|u_k^a - u^{\text{truth}}\|^2}{N}} \quad (25)$$

for  $u_k^a(\cdot)$  and

$$\sqrt{\frac{\sum_{k=1}^N \|u_k^a(0) - u^{\text{truth}}(0)\|^2}{N}} \quad (26)$$

for the initial state, where  $N$  is the number of numerical experiments. The RMSE of backgrounds,  $u_k^b(\cdot)$ , is defined in a similar way. In the Monte Carlo experiments, the backgrounds in eq. (21) are randomly generated so that the RMSE of  $u^b(\cdot)$  is 0.3638, which is 11.14% relative to the norm of the truth,  $u^{\text{truth}}(t)$ ; the RMSE of initial states  $u^a(0)$  equals 0.3524, which is 15.13% relative to the truth  $u_0$ . 4D-Var data assimilations are applied to all backgrounds in eq. (21) with sensor data in the form of eq. (22), from both the equally spaced sensor locations [eq. (19)] and the optimal locations [eq. (18)]. The data assimilation algorithm requires the background error covariance and the sensor noise covariance. They are generated from the set of errors  $\{u_k^b(0) - u_0 | k = 1, \dots, N_k = 200\}$  and noise  $(v_1, v_2, \dots, v_{N_t})$  normalised by  $N_t - 1$ .

The plots of the errors  $\|u^a(0) - u^{\text{truth}}(0)\|$  and  $\|u^a(\cdot) - u^{\text{truth}}(\cdot)\|$  from the Monte Carlo experiments are shown in Figs. 4 and 5. The estimation accuracy is significantly improved by using the optimal sensor locations. In fact, the RMSE of  $u_k^a(\cdot)$ ,  $k = 1, 2, \dots, 200$ , is 0.0788 for the equally spaced sensor locations. The RMSE is reduced to 0.0325 if the optimal sensor locations are used, which is an improvement of more than 59%. The errors of the initial states have similar behaviour. For equally spaced sensors, the RMSE of initial states is 0.1647; and the value is

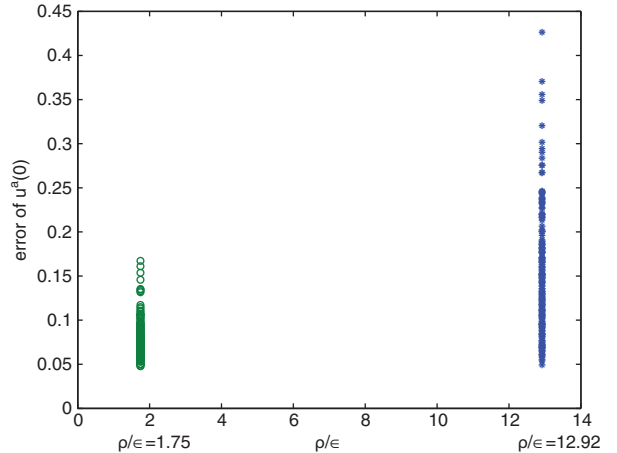


Fig. 4. The scatter plot of  $\rho/\epsilon$  vs.  $\|u^a(0) - u^{\text{truth}}(0)\|$ .



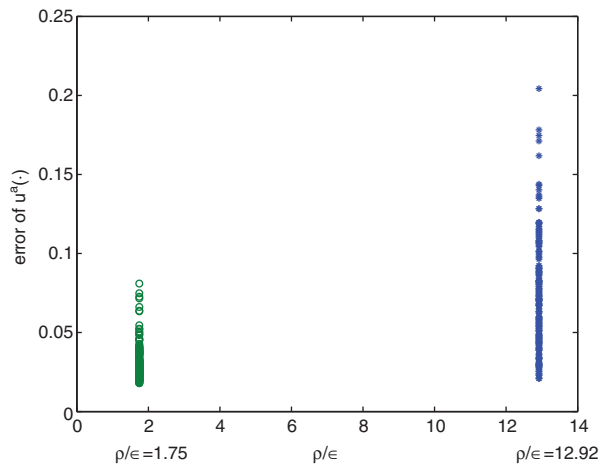


Fig. 5. The scatter plot of  $\rho/\epsilon$  vs.  $\|u^a(\cdot) - u^{\text{truth}}(\cdot)\|$ .

0.0786 for optimal sensor locations, which is a 52% error reduction. These results are summarised in Table 1.

The Monte Carlo experiments show that the concept of observability is consistent with the outcomes from data assimilation, i.e. data from sensors with higher observability results in higher estimation accuracy. This implies that sensor data with high observability contains more valuable information than those with low observability. However, one might notice from Table 1 that the scale of accuracy improvement does not match with the scale of the observability. In fact, the unobservability index of optimal sensor locations is 1.75 in contrast to 12.92 for equally spaced sensors, i.e. the observability is improved by about 86%. However, the improvement of estimation accuracy is about 50 ~ 60% depending on the choice of the norm. This phenomenon is largely due to the error covariance of initial background. In the computation of unobservability index, the impact of the error covariance of  $u_k^b(0)$  is not included in the cost function [eq. (16)]. However, one can certainly take this into consideration as a part of the cost function, i.e. a non-zero  $P_1$  in eq. (3). In the example of Burgers equation, we ignore the initial error covariance so that the observability is focused on the quality of sensor locations only. On the other hand, if the covariance of the background is known at the time of sensor location design, it can certainly be included in the cost function eq. (3) so that the optimal locations reflect the combined value of the information on both the initial error covariance and the sensor data.

Table 1. Summary of Monte Carlo experiments

	Unobservability index, $\rho/\epsilon$	RMSE of $u^a(0)$	RMSE of $u^a(\cdot)$
Equally spaced sensors	12.92	0.1647	0.0788
Optimal sensor location	1.75	0.0786	0.0325
Improvement	86%	52%	58%

Figure 6 shows the RMSE as a function of time. More specifically, using the 200 analysis resulting from data assimilation we define the following function

$$\sqrt{\frac{\sum_{k=1}^N \|u_k^a(t) - u^{\text{truth}}(t)\|^2}{N}}$$

where  $N=200$ . In Fig. 6, the curve represents the RMSE from equally spaced sensors and the dash curve represents the result using optimal sensor locations. Due to the non-linear nature of the system, the sensitivity of the state variables to sensor noise is not evenly distributed along a trajectory. Maximising observability results in an overall improvement of the estimation accuracy. More specifically, the optimal sensor locations result in significantly smaller error when the estimation is difficult and relatively inaccurate, which is in the time interval  $[0,1.3]$ . To compromise, the optimal sensor locations result in slightly bigger error in  $[1.3,5]$  when the estimate is relatively accurate.

The initial state and the interval of two times standard deviation are shown in Fig. 7, in which the curve is the result of using equally spaced sensors and the dash curve represents the result using optimal sensor locations, which has a significantly smaller error standard deviation.

## 6. Robustness analysis and some additional issues

In practical applications, robustness is an important factor in evaluating optimal designs. It is desired that suboptimal or acceptable results can be achieved in the presence of uncertainties. In data assimilations, the specification of the initial error covariance,  $\mathbf{P}_0^b$ , is a significant challenge due to its size and uncertainty. In the robustness study,  $\mathbf{P}_0^b$  is varied from its true value by about 10%. Then the RMSEs of  $u^a(\cdot)$  using equally spaced sensors and optimal locations are computed, which equal 0.0938 and 0.0351, respectively. If the covariance matrix is changed by about 50% from its true value, the RMSEs become 0.1191 and 0.0887. The results are summarised in Table 2. The robustness is also numerically tested for the sensor noise covariance,  $\mathbf{R}$ . If the matrix is changed by 100% from the true value, the RMSEs are 0.0646 and 0.0407 for equally spaced sensors and optimal sensor locations, respectively. In all cases, the optimal sensor locations result in significantly improved

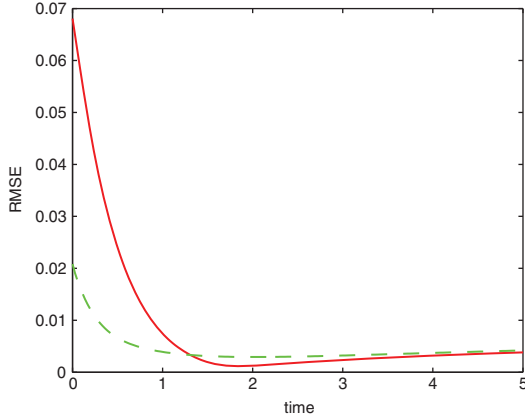


Fig. 6. RMSE as a function of time. Curve: equally spaced sensors. Dash: optimal sensor locations.

estimation accuracy, which ranges from 25% to 62% (see Table 2).

In addition to the variations of covariance matrices, another study is carried out to test the robustness under different probability distributions. In the previous numerical experiments, the background and sensor noise are both generated using normal distributions. To test the robustness, we generate a set of 200 backgrounds using uniform distribution in the interval of  $[-0.17, 0.17]$ . The sensor noise follows uniform distribution in  $[-0.003, 0.003]$ . Then 4D-Var data assimilation is applied to this set of data to generate two hundred  $u_k^a(\cdot)$ . The results are shown in Table 3. Consistent to the previous numerical results, the estimation error of  $u(0)$  is improved by 35% by using the optimal sensor locations. The improvement in the estimation of  $u(\cdot)$  is more than 42%.

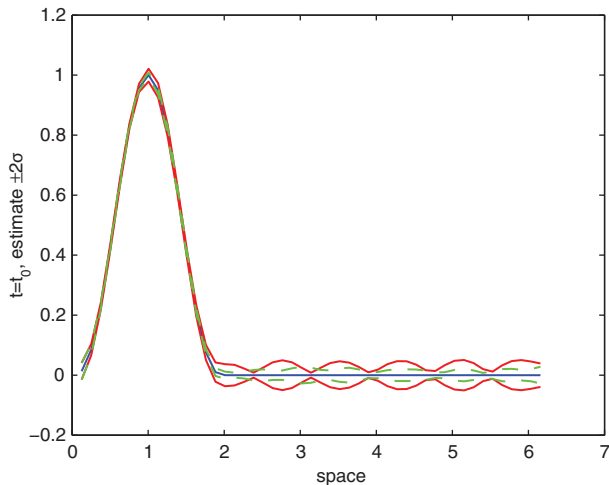


Fig. 7. The  $\pm 2\sigma$  interval around the initial state. Curve: equally spaced sensors. Dash: optimal sensor locations.

In the following, we revisit Definition 1 to address some related issues. What eq. (4) measures is, in fact, the observability of the initial state  $u_0$ . More specifically, the meaning of  $\varepsilon$  and  $\rho$  can be explained as follows: if the sensor error is bounded by  $\varepsilon$ , then the *worst* least square estimate  $u^a(0)$  of the initial state  $u_0$  has an error that is less than or equal to  $\rho$ . Because the trajectory of a system is uniquely determined by its initial state, for many systems an accurate estimate of  $u_0$  results in an accurate estimate of the entire trajectory. However, the initial state is not always the first priority in estimation. For instance, filtering problems require an accurate estimation of the current state, rather than the initial state. In the example of Burgers' equation, the system is stable in which the trajectory has active non-linear behaviour around the initial time. As a result, we focus on the observability of the initial state. Nevertheless, an accurate initial state does not automatically guarantee a good accuracy along the entire trajectory. In order to accommodate various requirements in estimation and prediction, it is necessary to define observability using different types of norms.

It is straightforward to generalise Definition 1 to the observability of  $u(k_0)$  for any fixed time  $k_0$ . Given any  $u(k_0) = u_0$ , it uniquely determines a trajectory  $u(k; u_0)$ . The associated cost,  $J(u_0, \delta u_0, \lambda)$ , can be defined in a similar way as in eq. (3). Let  $\rho > 0$  be a positive number. Then the number  $\varepsilon$  is defined as follows

$$\begin{aligned} \varepsilon^2 = \min_{\delta u_0} J(u_0, \delta u_0, \lambda) \\ \text{subject to} \\ \|\delta u_0\| = \rho \\ \delta u_0 \in W \end{aligned} \quad (27)$$

The ratio  $\rho/\varepsilon$  is a measure of the observability of  $u(k_0)$ . This number can be approximated using empirical Gramian matrix method introduced in Section 2, except that eq. (8) is defined using perturbations around  $u(k_0)$ .

Instead of focusing on a fixed point in the time interval, observability can be defined for an entire trajectory  $u(\cdot)$ . For this purpose, the norm of  $\delta u$  must measure the overall error along the trajectory, for instance the  $l^p$ -norm

$$\|\delta u\| = \left( \sum_{k=0}^{N_t} |\delta u(k)|^p \right)^{1/p}$$

Given a nominal trajectory  $u(k)$ , the space  $W$  is a subset of differentiable functions such that

$u(k) + \delta(k)$  is a trajectory that satisfies the dynamics model [eq. (1)]. For non-linear systems,  $W$  is not a linear space. It is a differential manifold with a metric. Although the space of differentiable functions is infinitely dimensional, the tangent space of  $W$  has a finite dimension. This is because trajectories are uniquely determined by initial

Table 2. Summary of a robustness study. The column under % represents the improvement in accuracy by using optimal sensor locations

	Variation of $P_0^b$	Variation of $\mathbf{R}$	RMSE of $u^a(\cdot)$	%
Equally spaced	10%	0	0.0938	
Optimal locations	10%	0	0.0351	62
Equally spaced	50%	0	0.1191	
Optimal locations	50%	0	0.0887	25
Equally spaced	0	100%	0.0646	
Optimal locations	0	100%	0.0407	37

Table 3. Summary of Monte Carlo experiments using uniform distributions

	Observability, $\rho/\varepsilon$	RMSE of $u^a(0)$	RMSE of $u^a(\cdot)$
Equally spaced sensors	12.92	0.1376	0.0645
Optimal sensor location	1.75	0.0893	0.0372
Improvement	86%	35%	42%

states, which are from a finite dimensional space. Suppose the output, i.e. the sensor measurement, is represented by

$$y(k) = h(k, u, (k))$$

The sensor measurement around the nominal trajectory is denoted by  $y(k; u + \delta u)$ , i.e.

$$y(k; u + \delta u) = h(k, u(k) + \delta u(k))$$

the variation of the measured variable  $y(k)$  around a nominal trajectory is defined as follows

$$J(u, \delta u) = \delta u(0)^T P_1 \delta u(0) + \|y(\cdot; u + \delta u) - y(\cdot; u)\|_Y \quad (28)$$

The observability of  $u(\cdot)$  is defined as follows,

$$\begin{aligned} \varepsilon^2 = \min_{\delta u} J(u, \delta u) \\ \text{subject to} \\ \|\delta u\| = \rho \\ \delta u \in W \end{aligned} \quad (29)$$

For this definition of observability, the empirical Gramian method can be modified to approximate  $\rho/\varepsilon$ .

## 7. Conclusions

To summarise, the concept of observability is defined as the theoretical foundation for the optimal design of sensor locations. The method of empirical Gramian matrix is a practical way to approximate the observability. It provides a cost function for the optimisation of sensor locations, which is carried out numerically using a gradient projection method. As a testbed, the Burgers equation is employed in the examples. The concept and computational methods are found applicable to these examples. The results prove that optimal sensor placement leads to significantly improved estimation accuracy in data assimilations when 4D-Var is used; and the performance is tested to be robust.

This work raises many questions for long term research. Optimal trajectory planning is an active research area in control systems in which the path, the control, and the trajectory of unmanned vehicles are designed to optimise a given performance measure (see, for instance, Tang et al. (2011) and references therein). The combination of algorithms in trajectory planning with the concept of observability is an obvious next step of research. The goal is to extend the results in this study on fixed sensor locations to mobile sensors in more dynamic environments. It requires optimal solutions with much higher dimensions that are subject to complicated constraints. In computation, developing numerical algorithms of observability for large scale systems is a fundamental challenge in meteorology applications. The examples in this study serve the purpose of proving concepts. However, the dimension of the problem is too low to verify the capability of finding optimal sensor design for high dimensional systems.

## 8. Acknowledgements

Supporting sponsors, Naval Research Laboratory and Air Force Office of Scientific Research, are gratefully acknowledged.

## References

- Baker, N. and Daley, R. 2000. Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. R. Meteorol. Soc.* **126**, 1431–1454.
- Baker, N. and Langland, R. 2009. Diagnostics for evaluating the impact of satellite observations. In: *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (eds. S. Park and L. Xu). Springer-Verlag, New York, 475 pp.
- Cardinali, C. 2009. Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.* **135**, 239–250.

- Cardinali, C., Pezzulli, S. and Andersson, E. 2004. Influence matrix diagnostics of a data assimilation system. *Q. J. R. Meteorol. Soc.* **130**, 2767–2786.
- Cardinali, C. and Prates, F. 2011. Performance measurement with advanced diagnostic tools of all-sky microwave imager radiances in 4D-Var. *Q. J. R. Meteorol. Soc.* **137**(661), 2038–2046. DOI:10.1002/qj.865.
- Daescu, D. 2008. On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. *Mon. Weather Rev.* **136**, 3050–3065.
- Daescu, D. and Todling, R. 2009. Adjoint estimation of the variation in model functional output due to the assimilation of data. *Mon. Weather Rev.* **137**, 1705–1716.
- Daescu, D. and Todling, R. 2010. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Q. J. R. Meteorol. Soc.* **136**, 2000–2012.
- Gelaro, R. and Zhu, Y. 2009. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*. **61A**, 179–193.
- Gelaro, R., Langland, R., Pellerin, S. and Todling, R. 2010. The THORPEX observation impact intercomparison experiment. *Mon. Weather Rev.* **138**, 4009–4025.
- Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations*. 3rd ed. The Johns Hopkins University Press, Baltimore, MD.
- Hogan, T. and Rosmond, T. 1991. The description of the navy operational global atmospheric prediction systems spectral forecast model. *Mon. Weather Rev.* **119**, 1786–1815.
- Kang, W. and Xu, L. 2009a. Computational analysis of control systems using dynamic optimization. *arXiv*, 0906.0215v2.
- Kang, W. and Xu, L. 2009b. A quantitative measure of observability and controllability. In: *IEEE Proceedings of Conference on Decision and Control*, Shanghai, China. 6413–6418.
- Krener, A.J. and Ide, K. 2009. Measures of unobservability. In: *IEEE Proceedings of Conference on Decision and Control*, Shanghai, China. 6401–6406.
- Langland, R. and Baker, N. 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*. **56A**, 189–201.
- Li, H., Kalnay, E. and Miyoshi, T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**, 523–533.
- Liu, J. and Kalnay, E. 2008. Estimating observation impact without adjoint model in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **134**, 1327–1335.
- Liu, J., Kalnay, E., Miyoshi, T. and Cardinali, C. 2009. Analysis sensitivity calculation in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**, 1842–1851.
- Polak, E. 1997. *Optimization: Algorithms and Consistent Approximations*. Springer, New York.
- Rosmond, T. 1997. *A Technical Description of the NRL Adjoint Modeling System*. Naval Research Laboratory Publication, NRL/MR/7532/97/7230, NRL, Monterey, CA.
- Tang, K., Wang, B., Kang, W. and Chen, B. M. 2011. Minimum time control of helicopter UAVs using computational dynamic optimization. In: *IEEE Proceedings of International Conference on Control & Automation*, Santiago, Chile. 848–852.
- Tremolet, Y. 2008. Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*. **60A**, 964–978.
- Xu, L. and Daley, R. 2000. Towards a true 4-dimensional data assimilation algorithm: application of a cycling representer algorithm to a simple transport problem. *Tellus*. **52A**, 109–128.
- Xu, L., Langland, R., Baker, N. and Rosmond, T. 2006. Development and testing of the adjoint of NAVDAS-AR. 9–13 October. In: *Proceedings of the Seventh International Workshop on Adjoint Applications in Dynamic Meteorology*, Obergurgl, Austria. 57–58.
- Xu, L., Rosmond, T. and Daley, R. 2005. Development of NAVDAS-AR: formulation and initial tests of the linear problem. *Tellus*. **57A**, 546–559.
- Xu, L., Rosmond, T., Goerss, J. and Chua, B. 2007. Toward a weak constraint operational 4D-Var system: application to the Burger’s equation. *Meteorologische Zeitschrift*. **16**(6), 1–13.