

On Controlling the Shape of the Cost Functional in Dynamic Data Assimilation: Guidelines for Placement of Observations and Application to Saltzman's Model of Convection

S. LAKSHMIVARAHAN

School of Computer Science, University of Oklahoma, Norman, Oklahoma

JOHN M. LEWIS

National Severe Storms Laboratory, Norman, Oklahoma, and Desert Research Institute, Reno, Nevada

JUNJUN HU

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

(Manuscript received 5 December 2019, in final form 1 June 2020)

ABSTRACT

Over the decades the role of observations in building and/or improving the fidelity of a model to a phenomenon is well documented in the meteorological literature. More recently adaptive/targeted observations have been routinely used to improve the quality of the analysis resulting from the fusion of data with models in a data assimilation scheme and the subsequent forecast. In this paper our goal is to develop an offline (preprocessing) diagnostic strategy for placing observations with a singular view to reduce the forecast error/innovation in the context of the classical 4D-Var. It is well known that the shape of the cost functional as measured by its gradient (also called adjoint gradient or sensitivity) in the control (initial condition and model parameters) space determines the marching of the control iterates toward a local minimum. These iterates can become marooned in regions of control space where the gradient is small. An open question is how to avoid these “flat” regions by bounding the norm of the gradient away from zero. We answer this question in two steps. We, for the first time, derive a linear transformation defined by a symmetric positive semidefinite (SPSD) Gramian $\mathbf{G} = \bar{\mathbf{F}}\bar{\mathbf{F}}$ that directly relates the control error to the adjoint gradient. It is then shown that by placing observations where the square of the Frobenius norm of $\bar{\mathbf{F}}$ (which is also the sum of the eigenvalues of \mathbf{G}) is a maximum, we can indeed bound the norm of the adjoint gradient away from zero.

1. Introduction

Judicious selection of the type (pressure, humidity, and wind, to name a few) and placement (the location within a given spatiotemporal domain of interest) of weather observations to advance understanding of a phenomenon and the concomitant practical ability to improve forecasts has a rich history in meteorology. Among the most celebrated use of weather observations to understand the ubiquitous extratropical cyclone was the fieldwork of Scandinavian meteorologists Jacob Bjerknes and Eric Palmén (Bjerknes and Palmén 1937). They organized the simultaneous release of 120 radiosondes from 11

European countries over a 2-day period that led to diagnosis of motions in the upper troposphere along a 3500-mi (5630-km) latitudinal cross section. Further, the study made it clear that the meridional gradient of the Coriolis parameter was an important parameter in the dynamics of upper waves. This revelation inspired Carl-Gustaf Rossby to derive his celebrated wave formula (Eliassen 1995).

In the spirit of the Bjerknes–Palmén study, but in our current day, the use of research aircraft to obtain thermodynamic and wind structure in the vicinity of Atlantic hurricanes through use of Omega dropwindsondes (ODWs) has gone far to improve the track forecasting of hurricanes (Burpee et al. 1996). Transmission of these data to operational prediction centers in real time permitted

Corresponding author: S. Lakshmivarahan, varahan@ou.edu

DOI: 10.1175/JAS-D-19-0329.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

inclusion of these data into the operational models. The improvement in track forecasting as tested on numerous hurricanes was impressive—track error reductions from 16% to 30%. Also impressive was the hurricane forecasting experience of the Burpee team that led them to choose observation locations on the hurricane’s periphery as well as near the storm’s center where the thermodynamic and wind data were critically important as input to track forecasting. Majumdar (2016) and Majumdar et al. (2011) present an excellent review of “targeted observations,” which classified the Burpee et al. (1996) as an example of targeting based on synoptic reasoning—not only an appropriate label for this contribution but also for the earlier contribution by Bjerknes and Palmén (1937).

Starting from the mid-1990s there has been a growing body of work within the meteorological literature related to (i) the analysis of impact of initial condition and observation on a chosen scalar metric related to forecast quality (over a subdomain or the entire domain of interest) and (ii) targeted/adaptive observation aimed at finding the type of a new set of observations that will bring improvement to forecast quality, such as, reduction of analysis covariance, and (iii) when there is a large amount of correlated set of observations (as may happen with satellites or radars), there has been great interest in the so-called “thinning” and/or creating superobservations by computing, say, an average over a given subdomain and associating that observation with the centroid of the subdomain.

The techniques used in these areas may be broadly divided into two groups: either based on the adjoint approach rooted in the classical variational analysis or using one of the many ensemble implementations of Kalman filtering approach to data assimilation. While much of the literature concentrates on the choice of the suitable type of observations to improve the forecast quality, a few also consider distribution of observation to answer a similar goal.

Specifically, Berliner et al. (1999) examine the application of statistical experimental design (Cochran and Cox 1992) to adaptively decide on supplementary observations that would improve the analyses. Torn and Hakim (2008), Ancell and Hakim (2007, hereafter referred to as AH2007), and Hakim and Torn (2008) use ensemble methods for targeting observations. AH2007 also contains a comparison of the adjoint based and ensemble based approaches to observation targeting. Lorenz and Emanuel (1998), Langland et al. (1999), Langland and Baker (2004), Evans et al. (1998), Koch et al. (2018), Manohar et al. (2018), Kotsuki et al. (2019), Majumdar et al. (2011), and Majumdar (2016)

contains extensive discussion of “thinning,” creating superobservations, targeted-observation strategies including adjoint sensitivity, forecast sensitivity to observations, ensemble transform filter method, and ensemble sensitivity. All these strategies are designed to improve analyses and ultimately reduce forecast error.

Our approach falls outside the strategies used in targeted/adaptive observations although the goal has strong overlap with those methodologies. It is assumed that we have observations of a given type represented by a vector \mathbf{Z} in the observation space R^m . It is well known that when the model map \mathbf{M} and/or the forward operator \mathbf{h} (that relates the observation to the state) are nonlinear, the cost functional may exhibit multiple minima (Lewis et al. 2006, example 24.4.1 in chapter 24). In addition to these intrinsic multiple minima that are purely dependent on the given nonlinearities in \mathbf{M} and \mathbf{h} , and the covariance matrix of the observational noise, the shape of the cost functional in the control space (consisting of initial conditions and model parameters) is also critically dependent on distribution (location) of the observations in the state space. Our goal is to control the shape of the cost functional by analyzing the dependence of its gradient (with respect to the control) on the (spatiotemporal) location of the observations. Namely, force the magnitude of this gradient to be bounded away from zero outside the intrinsic local minimum.

This is achieved through a strategy that decides on observation placement in space and time but in the context of classical four-dimensional variational data assimilation (4D-Var) (LeDimet and Talagrand 1986; Lewis and Derber 1985). It is well known that the shape of the cost function as measured by its gradient in control space determines the marching of the control iterates toward a local minimum of the cost function. The iterates of the minimization algorithm can become marooned in regions where the gradient is very small. An open question is how to avoid these “flat” regions in control space by guaranteeing that the norm of the gradient of the cost function is bounded away from zero. While much of the earlier work deal with sensitivity of the cost functional only with respect to the initial conditions alone, our analysis simultaneously deals with initial conditions and parameters on the same footing.

This question is answered in two steps. First, we derive a linear transformation defined by a symmetric positive semidefinite (SPSD) Gramian¹ $\mathbf{G} = \mathbf{F}^T \mathbf{F}$ that maps the error in the control to the gradient of the cost

¹ If $\mathbf{A} \in R^{m \times n}$, then the symmetric matrices $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ are known as Gramian of \mathbf{A} .

functional alias adjoint sensitivity or gradient,² where the component matrix $\bar{\mathbf{F}}$ is the product of 1) the forward sensitivity of the model solution, 2) the Jacobian of the forward operator and 3) the square root of the inverse of the covariance matrix of the observation noise. Refer to Fig. 1 for illustration. It is then shown that by placing observations where the square of the Frobenius norm of $\bar{\mathbf{F}}$ attains a maximum, we can indeed bound the norm of the adjoint gradient away from zero.

In section 2, we provide a short summary of the forecast error correction problem by relating the adjoint sensitivity or adjoint gradient to the innovation or forecast error. By relating this forecast error to the model’s forward sensitivity, we derive our main results in section 3. The intrinsic relation that exists between the adjoint sensitivity and the control error is further analyzed in section 4. The impact of observation placement on controlling the shape of the cost functional is analyzed in section 5. Two examples are considered in section 6: an air–sea interaction problem in time alone, and Saltzman’s (1962) more complicated nonlinear model of convection in space and time. Both examples exhibit the methodology’s avoidance of flat gradients in control space. Section 7 contains recommendations for placement of observations and concluding remarks. Appendix A contains mathematical details that are key to the analysis in sections 3–5. The Gramian structure of \mathbf{G} is developed in appendix B. In the short appendix C, we further illustrate the difference between our approach and a typical adjoint sensitivity analysis in AH2007.

2. Forecast error correction problem

a. Model equations

Let $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T \in R^n$ be the state of a nonlinear, deterministic, discrete time dynamic model at times $k = 0, 1, 2, \dots$. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)^T \in R^p$ be the vector of p model parameters and let $\mathbf{M}: R^n \times R^p \rightarrow R^n$ denote the (one-step state transition) model map where $\mathbf{M}(\mathbf{x}, \boldsymbol{\alpha}) = [M_1(\mathbf{x}, \boldsymbol{\alpha}), M_2(\mathbf{x}, \boldsymbol{\alpha}), \dots, M_n(\mathbf{x}, \boldsymbol{\alpha})]^T \in R^n$. Let

$$\mathbf{x}(k+1) = \mathbf{M}[\mathbf{x}(k), \boldsymbol{\alpha}] \tag{2.1}$$

be the given dynamical system whose solution starting from an initial state $\mathbf{x}(0)$ denotes the model forecast. Let $\mathbf{x}(k) = \mathbf{x}[k, \mathbf{x}(0), \boldsymbol{\alpha}]$ denote the solution of (2.1).

²The term adjoint sensitivity or gradient derives its name from the adjoint method used to compute it. Refer to Lewis et al. (2006, chapters 22–25) for details.

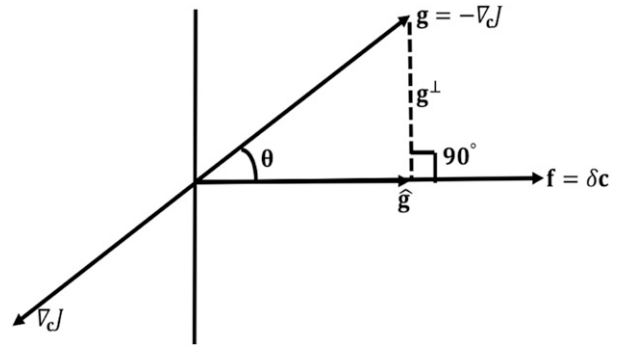


FIG. 1. A pictorial view of the linear relation $\mathbf{g} = \mathbf{G}\mathbf{f}$ between $\mathbf{f} = \delta\mathbf{c}$, the error in the initial in control \mathbf{c} and $\mathbf{g} = -\nabla_c J$ the negative of the gradient of the cost function $J(\mathbf{c})$, where $\mathbf{G} = \bar{\mathbf{F}}^T \bar{\mathbf{F}}$. The component $\hat{\mathbf{g}}$ is the orthogonal projection of \mathbf{g} along \mathbf{f} . That is, $\hat{\mathbf{g}} = \hat{\mathbf{f}}(\mathbf{f}, \mathbf{g}) = \mathbf{P}_f \mathbf{g}$, where $\hat{\mathbf{f}}$ is the unit vector along \mathbf{f} and \mathbf{P}_f is the orthogonal projection matrix along \mathbf{f} .

Since the pair $\mathbf{c} = [\mathbf{x}^T(0), \boldsymbol{\alpha}^T]^T \in R^n \times R^p$ controls the evolution of $\mathbf{x}(k)$, it is known as the control. It is tacitly assumed that the model in (2.1) is the best known model for the process in question and is suitably parameterized to capture the inherent natural variations. Clearly, simultaneous estimation of the initial conditions and parameters in dynamic data assimilation puts this par with the goals and methods used in the “adaptive control” literature (Narendra and Annaswamy 2005).

b. Observation

Let $\mathbf{z}(k) \in R^m$ be the vector observation at time k given by

$$\mathbf{z}(k) = \mathbf{h}[\bar{\mathbf{x}}(k) + \boldsymbol{\xi}(k)], \tag{2.2}$$

where $\mathbf{h}: R^n \rightarrow R^m$ and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_m(\mathbf{x})]^T \in R^m$ is called the forward operator that relates the model space R^n and the observation space R^m ; $\bar{\mathbf{x}}(k)$ is the true but unknown state of nature at time k and $\boldsymbol{\xi}(k) \sim N[0, \mathbf{R}(k)]$ is a zero mean, temporally uncorrelated, Gaussian noise with $\mathbf{R}(k) \in R^{m \times m}$ as its covariance matrix at time k . It is assumed that $\mathbf{R}(k)$ is known and positive definite in addition to being symmetric. In the case that the forward operator is linear, (2.2) takes the form $\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \boldsymbol{\xi}(k)$ where $\mathbf{H} \in R^{m \times n}$ is a matrix that takes model output and converts it to the model’s counterpart of the observations.

c. Assumption

It is assumed that the model map \mathbf{M} in (2.1), the forward operator \mathbf{h} in (2.2) and the model solution $\mathbf{x}(k)$ are continuously differentiable in their arguments. That is, the following Jacobians are well defined for all $\mathbf{x}, \boldsymbol{\alpha}$, and $k \geq 0$:

$$\begin{aligned} \mathbf{D}_M(k) &= \left[\frac{\partial M_i(\mathbf{x}, \boldsymbol{\alpha})}{\partial x_j} \right]_{\mathbf{x}=\mathbf{x}(k)} \in \mathbf{R}^{n \times n}, \\ \mathbf{D}_M^\alpha(k) &= \left[\frac{\partial M_i(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_j} \right]_{\mathbf{x}=\mathbf{x}(k)} \in \mathbf{R}^{n \times p}, \\ \mathbf{D}_h(k) &= \left[\frac{\partial h_i(\mathbf{x})}{\partial x_j} \right]_{\mathbf{x}=\mathbf{x}(k)} \in \mathbf{R}^{n \times m}, \end{aligned} \quad (2.3a)$$

$$\begin{aligned} \mathbf{U}(k) &= [U_{ij}(k)] = \left[\frac{\partial x_i(k)}{\partial x_j(0)} \right] \in \mathbf{R}^{n \times n} \quad \text{and} \\ \mathbf{V}(k) &= [V_{ij}(k)] = \left[\frac{\partial x_i(k)}{\partial \alpha_j} \right] \in \mathbf{R}^{n \times p}. \end{aligned} \quad (2.3b)$$

d. Innovation/forecast error

Let $\mathbf{z}(k)$ be a single observation given at time k . The difference

$$\bar{\mathbf{e}}(k) = \mathbf{z}(k) - \mathbf{h}[\mathbf{x}(k)] \quad (2.4)$$

is called the innovation or the forecast error. Refer to Fig. 2 for an illustration.

e. A classification of forecast errors

For the purposes of our analysis in this paper, it is useful to classify the forecast errors into three types (Lakshmivaran and Lewis 2010; Lakshmivaran et al. 2017). Type 1 error is said to occur when the model is assumed to be perfect and the true control $\bar{\mathbf{c}} = [\bar{\mathbf{x}}^T(0), \bar{\boldsymbol{\alpha}}^T]^T$ is known. In this case, $\mathbf{x}(k) = \bar{\mathbf{x}}(k)$ and the forecast error is temporally uncorrelated and is a pure Gaussian noise that cannot be controlled. Type 2 error occurs when the model is perfect, but the control \mathbf{c} differs from the true but unknown control $\bar{\mathbf{c}}$. In this case, the forecast error $\bar{\mathbf{e}}(k)$ depends only on the control error, $\delta \mathbf{c} = [\delta \mathbf{x}^T(0), \delta \boldsymbol{\alpha}^T]^T$ and

$$\delta \mathbf{x}(k) = \bar{\mathbf{x}}(k) - \mathbf{x}(k) \quad \text{for } k \geq 0, \quad \delta \boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}. \quad (2.5)$$

Type 3 error is said to occur when the model is not perfect and the true control is not known. In this case, the forecast error is confounded by the unknown model error and the unknown control error. In this paper, we confine our analysis to the simpler case of type 2 errors. Estimation of type 3 errors is contained in Lakshmivaran et al. (2013).

f. Statement of the problem

The forecast error correction problem is usually recast as the minimization of the weighted sum of squared

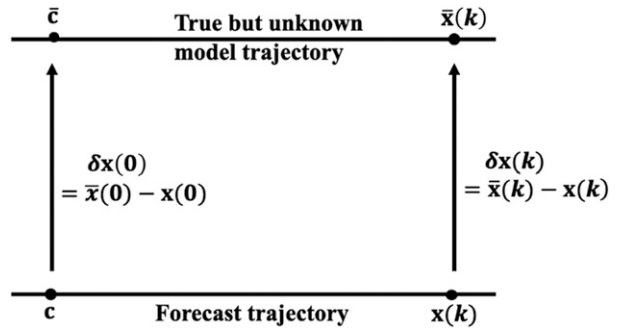


FIG. 2. $\bar{\mathbf{x}}(k)$ is the true but unknown state $\mathbf{x}(k)$ of the model at time k starting from the true but unknown control $\bar{\mathbf{c}} = [\bar{\mathbf{x}}^T(0), \bar{\boldsymbol{\alpha}}^T]^T$. $\mathbf{x}(k)$ is the model forecast at time k starting from the erroneous initial control $\mathbf{c} = [\mathbf{x}^T(0), \boldsymbol{\alpha}^T]^T \neq \bar{\mathbf{c}}$. $\delta \mathbf{x}(k) = \bar{\mathbf{x}}(k) - \mathbf{x}(k)$ is the first variation in the model state at time k indicated by the first variation $\delta \mathbf{c} = [\delta \mathbf{x}^T(0), \delta \boldsymbol{\alpha}^T]^T$, where $\delta \mathbf{x}(0) = \bar{\mathbf{x}}(0) - \mathbf{x}(0)$ and $\delta \boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$. $\mathbf{z}(k)$ is the observation at time k that contains information on $\bar{\mathbf{x}}(k)$ through the relation $\mathbf{z}(k) = \mathbf{h}[\bar{\mathbf{x}}(k)]$. The forecast error $\bar{\mathbf{e}}(k) = \mathbf{z}(k) - \mathbf{h}[\mathbf{x}(k)]$.

error criterion. To this end, define a cost functional $J: \mathbf{R}^n \times \mathbf{R}^p \rightarrow \mathbf{R}$ given by

$$J(\mathbf{c}) = \|\bar{\mathbf{e}}(k)\|_{\mathbf{R}^{-1}(k)}^2 = \frac{1}{2}(\bar{\mathbf{e}}(k), \mathbf{R}^{-1}(k)\bar{\mathbf{e}}(k)), \quad (2.6)$$

where $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the standard inner product. Again, we hasten to add that if prior information on either the initial condition $\mathbf{X}(0)$ or the parameter $\boldsymbol{\alpha}$ or both is available, we could add a quadratic penalty term to the cost function, which will contribute another term in the expression for adjoint sensitivity.

Since $\bar{\mathbf{e}}(k)$ depends on the forecast $\mathbf{x}(k)$ that in turn is a function of the control \mathbf{c} , the cost functional is an implicit function of \mathbf{c} through the given model Eq. (2.1). We hasten to add that since the cost functional is additive in the number of observations, without loss of generality, it is assumed that there is only one observation $\mathbf{z}(k) \in \mathbf{R}^m$ at time k . This constrained minimization problem is usually solved as a strong constraint problem using the standard Lagrangian multiplier method. The resulting framework has come to be known as the adjoint method (see Lewis et al. 2006, chapters 22–24).

3. Fine structure of adjoint sensitivity

Instead of determining the adjoint sensitivity through use of the tangent linear model as typically done in the standard 4D-Var mechanics (see Lewis et al. 2006, chapter 24), we derive an alternate expression for $\delta \mathbf{x}(k)$ in terms of the forecast sensitivity to control and the associated incremental change in control.

a. *Alternate expression for $\delta\mathbf{x}(k)$*

The first-order Taylor expansion for $\delta\mathbf{x}(k)$ is given by

$$\delta\mathbf{x}(k) = \mathbf{U}(k)\delta\mathbf{x}(0) + \mathbf{V}(k)\delta\boldsymbol{\alpha}, \quad (3.1)$$

where the dynamics of evolution of the forward sensitivity matrices $\mathbf{U}(k)$ and $\mathbf{V}(k)$ defined in (2.3b) are given in the following (see Lakshmivarahan et al. 2017, chapters 2 and 3):

$$\begin{aligned} \mathbf{U}(k+1) &= \mathbf{D}_m(k)\mathbf{U}(k), \quad \text{with initial condition,} \\ \mathbf{U}(0) &= \mathbf{I}, \end{aligned} \quad (3.2)$$

$$\begin{aligned} \mathbf{V}(k+1) &= \mathbf{D}_m(k)\mathbf{V}(k) + \mathbf{D}_m^a(k), \quad \text{with initial condition} \\ \mathbf{V}(0) &= \mathbf{0}. \end{aligned} \quad (3.3)$$

b. *An alternate expression for the forecast error $\bar{\mathbf{e}}(k)$*

Using (2.2), we get an expression for the actual forecast error as

$$\bar{\mathbf{e}}(k) = \mathbf{z}(k) - \mathbf{h}[\mathbf{x}(k)] = \mathbf{h}[\bar{\mathbf{x}}(k)] - \mathbf{h}[\mathbf{x}(k)] + \boldsymbol{\xi}(k). \quad (3.4)$$

But using the expression for $\delta\mathbf{x}(k)$ in (2.5), a first-order approximation $\mathbf{e}(k)$ to $\bar{\mathbf{e}}(k)$ is given by

$$\begin{aligned} \mathbf{e}(k) &= \mathbf{h}[\mathbf{x}(k) + \delta\mathbf{x}(k)] - \mathbf{h}[\mathbf{x}(k)] + \boldsymbol{\xi}(k) \\ &= \mathbf{D}_h(k)\delta\mathbf{x}(k) + \boldsymbol{\xi}(k), \end{aligned} \quad (3.5)$$

where the Jacobian $\mathbf{D}_h(k) \in R^{m \times n}$ is given in (2.3a). Further, substituting (3.1) in (3.5) and regrouping, we get an alternate expression for $\mathbf{e}(k)$ as

$$\mathbf{e}(k) = \mathbf{A}(k)\delta\mathbf{x}(0) + \mathbf{B}(k)\delta\boldsymbol{\alpha} + \boldsymbol{\xi}(k), \quad (3.6a)$$

where $\mathbf{A}(k) = \mathbf{D}_h(k)\mathbf{U}(k)$, $\mathbf{B}(k) = \mathbf{D}_h(k)\mathbf{V}(k)$.

Indeed, the correction $\delta\mathbf{c}$ to the control \mathbf{c} needed to reduce the forecast error (modulo observation noise) can be readily obtained by solving a linear least squares problem (Lakshmivarahan et al. 2017, chapter 2) obtained by rewriting (3.6a) as

$$\mathbf{e}(k) = \mathbf{S}(k)\delta\mathbf{c} + \boldsymbol{\xi}(k), \quad (3.6b)$$

where $\mathbf{S}(k) = [\mathbf{A}(k), \mathbf{B}(k)] \in R^{m \times (n+p)}$.

Remark 3.1: In the 4D-Var framework, the adjoint sensitivity is computed using the actual forecast error, $\bar{\mathbf{e}}(k)$ in (3.4). Refer to chapter 24 in Lewis et al. (2006) for details. However, in this alternate approach, we used the first-order approximation $\mathbf{e}(k)$ in (3.5). When $\mathbf{h}(\mathbf{x})$ is linear, that is, $\mathbf{h}(\mathbf{x}) = \mathbf{H}\mathbf{x}$, then $\bar{\mathbf{e}}(k) = \mathbf{e}(k)$. This alternate method for quantifying the forecast error is known as the first-order forward sensitivity method (FSM) and is closely related to the 4D-Var formulation. For details

refer to Lakshmivarahan and Lewis (2010) and the monograph Lakshmivarahan et al. (2017).

c. *Fine structure of adjoint sensitivity*

It can be verified that the exact first variation δJ_e in $J(\mathbf{c})$ resulting from the initial variation $\delta\mathbf{c}$ in \mathbf{c} is given by (see Lewis et al. 2006, chapter 24)

$$\delta J_e = -\langle \mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\bar{\mathbf{e}}(k), \delta\mathbf{x}(k) \rangle. \quad (3.7)$$

Replacing $\bar{\mathbf{e}}(k)$ in (3.7) by $\mathbf{e}(k)$, we get an approximation δJ to δJ_e given by

$$\delta J = -\langle \mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\mathbf{e}(k), \delta\mathbf{x}(k) \rangle. \quad (3.8)$$

Substituting $\mathbf{e}(k)$, $\mathbf{S}(k)$, $\mathbf{A}(k)$, and $\mathbf{B}(k)$ from (3.6)–(3.7) and $\delta\mathbf{x}(k)$ from (3.1) and regrouping, we express δJ as the sum of the deterministic component, δJ^d and the stochastic component δJ^n , which are given as follows:

$$\begin{aligned} \delta J^d &= -\langle \bar{\mathbf{H}}(k)[\mathbf{U}(k)\delta\mathbf{x}(0) + \mathbf{V}(k)\delta\boldsymbol{\alpha}], \\ &\quad [\mathbf{U}(k)\delta\mathbf{x}(k) + \mathbf{V}(k)\delta\boldsymbol{\alpha}], \end{aligned} \quad (3.9a)$$

$$\delta J^n = -\langle \mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\boldsymbol{\xi}(k), [\mathbf{U}(k)\delta\mathbf{x}(0) + \mathbf{V}(k)\delta\boldsymbol{\alpha}] \rangle. \quad (3.9b)$$

Define a symmetric matrix

$$\bar{\mathbf{H}}(k) = \mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\mathbf{D}_h(k) \in R^{n \times n} \quad (3.10)$$

that depends only on the observation system. Again, collecting the like terms and simplifying, it is immediate that

$$\begin{aligned} \delta J^d &= -\langle \mathbf{U}^T(k)\bar{\mathbf{H}}[\mathbf{U}(k)\delta\mathbf{x}(0) + \mathbf{V}(k)\delta\boldsymbol{\alpha}], \delta\mathbf{x}(0) \rangle \\ &\quad - \langle \mathbf{V}^T(k)\bar{\mathbf{H}}[\mathbf{U}(k)\delta\mathbf{x}(0) + \mathbf{V}(k)\delta\boldsymbol{\alpha}], \delta\boldsymbol{\alpha} \rangle \end{aligned} \quad (3.11a)$$

and

$$\begin{aligned} \delta J^n &= -\mathbf{U}^T(k)\mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\boldsymbol{\xi}(k), \delta\mathbf{x}(0) \\ &\quad - \langle \mathbf{V}^T(k)\mathbf{D}_h^T(k)\mathbf{R}^{-1}(k)\boldsymbol{\xi}(k), \delta\boldsymbol{\alpha} \rangle. \end{aligned} \quad (3.11b)$$

From first principles,³ we can readily identify that the gradients of J :

$$\nabla_{\mathbf{x}(0)} J(\mathbf{c}) = -[\mathbf{U}^T\bar{\mathbf{H}}\mathbf{U}]\delta\mathbf{x}(0) + [\mathbf{U}^T\bar{\mathbf{H}}\mathbf{V}]\delta\boldsymbol{\alpha} - \mathbf{U}^T\mathbf{H}\boldsymbol{\xi}(k), \quad (3.12a)$$

³ If $f: R^n \rightarrow R$ is continuously differentiable, then $\delta f = \nabla f \cdot \delta\mathbf{x}$ (Lewis et al. 2006, appendix C).

$$\nabla_{\mathbf{c}} J(\mathbf{c}) = -[\mathbf{V}^T \overline{\mathbf{H}} \mathbf{U}] \delta \mathbf{x}(0) + [\mathbf{V}^T \overline{\mathbf{H}} \mathbf{V}] \delta \mathbf{c} - \mathbf{V}^T \underline{\mathbf{H}} \boldsymbol{\xi}(k), \tag{3.12b}$$

where

$$\underline{\mathbf{H}}(k) = \mathbf{D}_h^T(k) \mathbf{R}^{-1}(k) \in R^{n \times m}. \tag{3.13}$$

These relations clearly bring out the impact of interaction resulting from simultaneous variation of both the initial condition and the parameters and the observation noise covariance on the adjoint sensitivities.

Setting

$$\nabla_{\mathbf{c}} J = \begin{bmatrix} \nabla_{\mathbf{x}(0)} J \\ \nabla_{\mathbf{c}} J \end{bmatrix} \quad \text{and} \quad \delta \mathbf{c} = \begin{bmatrix} \delta \mathbf{x}(0) \\ \delta \mathbf{c} \end{bmatrix},$$

(3.11)–(3.13) can be succinctly expressed as

$$\begin{aligned} \nabla_{\mathbf{c}} J &= \nabla_{\mathbf{c}} J^d + \nabla_{\mathbf{c}} J^n, \quad \nabla_{\mathbf{c}} J^d = -[\mathbf{F}^T \overline{\mathbf{H}} \mathbf{F}] \delta \mathbf{c} \quad \text{and} \\ \nabla_{\mathbf{c}} J^n &= -\mathbf{F}^T \underline{\mathbf{H}}(k) \boldsymbol{\xi}(k), \end{aligned} \tag{3.14}$$

where

$$\mathbf{F} = [\mathbf{U}, \mathbf{V}] \in R^{n \times (n+p)}, \quad \mathbf{F}^T = \begin{bmatrix} \mathbf{U}^T \\ \mathbf{V}^T \end{bmatrix} \in R^{(n+p) \times n} \tag{3.15}$$

and

$$\begin{aligned} \mathbf{F}^T \overline{\mathbf{H}} \mathbf{F} &= \begin{bmatrix} \mathbf{U}^T \\ \mathbf{V}^T \end{bmatrix} \overline{\mathbf{H}} [\mathbf{U}, \mathbf{V}] \\ &= \begin{bmatrix} \mathbf{U}^T \overline{\mathbf{H}} \mathbf{U} & \mathbf{U}^T \overline{\mathbf{H}} \mathbf{V} \\ \mathbf{V}^T \overline{\mathbf{H}} \mathbf{U} & \mathbf{V}^T \overline{\mathbf{H}} \mathbf{V} \end{bmatrix} \in R^{(n+p) \times (n+p)} \end{aligned} \tag{3.16}$$

is the outer product matrix, which inherits the Gramian structure alluded to in the introduction. Stated in other words, there exists a linear relation between $\delta \mathbf{c}$, the error in the control to the deterministic part of the adjoint gradient, $\nabla_{\mathbf{c}} J^d$, and a linear relation between the observation noise $\boldsymbol{\xi}(k)$ and the stochastic component $\nabla_{\mathbf{c}} J^n$. This new expression for the adjoint sensitivity clearly brings out (i) the role of the model through the forward sensitivities \mathbf{U} and \mathbf{V} on one hand, and (ii) the role of the observation system through $\overline{\mathbf{H}} = \mathbf{D}_h^T \mathbf{R}^{-1} \mathbf{D}_h$ on the other. Indeed, this separability of the effect of the model and the observation system clearly provides the “fine structure” of adjoint sensitivity that, to our knowledge, is hitherto unknown in the literature.

Remark 3.2: The standard assumption about the observation noise $\boldsymbol{\xi}(k)$ is that it is zero mean, temporally uncorrelated, Gaussian with $\mathbf{R}(k)$ as its covariance at

time k ; that is, $\boldsymbol{\xi}(k) \sim N[0, \mathbf{R}(k)]$. Hence, from (3.14), the expected value and covariance (Cov) are given by

$$E(\nabla_{\mathbf{c}} J) = \nabla_{\mathbf{c}} J^d \tag{3.17}$$

and

$$\begin{aligned} \text{Cov}(\nabla_{\mathbf{c}} \mathbf{J}) &= E[(\nabla_{\mathbf{c}} \mathbf{J}^n)(\nabla_{\mathbf{c}} \mathbf{J}^n)^T] \\ &= \mathbf{F}^T \underline{\mathbf{H}} \mathbf{R} \underline{\mathbf{H}}^T \mathbf{F} = \mathbf{F}^T \mathbf{D}_h^T \mathbf{R}^{-1}(k) \mathbf{D}_h \mathbf{F} \\ &= \mathbf{F}^T \overline{\mathbf{H}} \mathbf{F} \end{aligned} \tag{3.18}$$

using (3.10) and (3.13). Henceforth, following the tradition in statistical physics (Tolman 2010), we invoke the principle of mean field analysis and concentrate only on the deterministic component $\nabla_{\mathbf{c}} J^d$ of the adjoint gradient. For simplicity in the notation, in the following, we identify $\nabla_{\mathbf{c}} J$ with $\nabla_{\mathbf{c}} J^d$. Further, appendix C clearly brings out the difference between the approach of this paper and a typical adjoint sensitivity analysis AH2007.

4. A tale of two vectors: $\delta \mathbf{c}$ and $\nabla_{\mathbf{c}} J$ —Further analysis

An immediate import of the above analysis is the deviation of the simple but key relation in (3.14) where the matrix $\mathbf{F}^T \overline{\mathbf{H}} \mathbf{F}$, that maps the control space $R^{(n+p)}$ of unknown control error vector $\delta \mathbf{c}$ into the space $R^{(n+p)}$ of the adjoint sensitivity $\nabla_{\mathbf{c}} J$.

As a prelude to further analysis, let us further simplify the notation by defining the following equivalences:

$$\mathbf{g} = -\nabla_{\mathbf{c}} J, \quad \mathbf{f} = \delta \mathbf{c} \quad \text{and} \quad \mathbf{G} = \mathbf{F}^T \overline{\mathbf{H}} \mathbf{F}. \tag{4.1}$$

Consequently, from (3.14) we obtain a fundamental relation

$$\mathbf{g} = \mathbf{G} \mathbf{f}. \tag{4.2}$$

Remark 4.1: An estimate of the Hessian of $J(\mathbf{c})$ at the minimum: It is well known that if \mathbf{A} is an SPD matrix and \mathbf{b} is a vector, then $\nabla \mathbf{Q}(\mathbf{x}) = \mathbf{A} \mathbf{x} - \mathbf{b}$ if and only if $\mathbf{Q}(\mathbf{x}) = (1/2) \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} + d$ for some arbitrary constant d (Meyer 2000; Lewis et al. 2006, chapter 10). From (4.1) and (4.2), we get $\nabla_{\mathbf{c}} J = -\mathbf{G} \mathbf{f} = -\mathbf{G}(\bar{\mathbf{c}} - \mathbf{c}) = \mathbf{G} \mathbf{c} - \mathbf{b}$ where $\mathbf{b} = \mathbf{G} \bar{\mathbf{c}}$. Combining these two facts, we can locally approximate $J(\mathbf{c})$ around $\bar{\mathbf{c}}$ by a quadratic functional given by $\bar{J}(\mathbf{c}) = (1/2) \mathbf{c}^T \mathbf{G} \mathbf{c} - \mathbf{b}^T \mathbf{c}$. Hence, the Hessian of $\bar{J}(\mathbf{c})$ is given by $\nabla^2 \bar{J}(\mathbf{c}) \approx \nabla^2 J(\mathbf{c}) = \mathbf{G} = \mathbf{F}^T \overline{\mathbf{H}} \mathbf{F}$, which by (3.18) is also the covariance of $\nabla_{\mathbf{c}} J$. Clearly, this Hessian is independent of the observation and its condition number controls the shape of $J(\mathbf{c})$.

The meaning and the power of this new relation in (4.2) is captured in Fig. 1 from which it is evident that we can express \mathbf{g} as the sum two orthogonal components:

$$\mathbf{g} = \hat{\mathbf{g}} + \hat{\mathbf{g}}^\perp, \tag{4.3}$$

where $\hat{\mathbf{g}}$ is the orthogonal projection of \mathbf{g} along \mathbf{f} and $\hat{\mathbf{g}}^\perp = \mathbf{g} - \hat{\mathbf{g}}$.

Thus, when we use \mathbf{g} routinely in the minimization of cost functional in the 4D-Var method (Lewis et al. 2006, chapters 22–24), it is now clear from Fig. 1 that it is not \mathbf{g} , but only its component $\hat{\mathbf{g}}$ that is responsible for bringing the reduction in the value of the cost functional and hence the forecast error. An important question is: what factors affect the magnitude of $\hat{\mathbf{g}}$? The 4D-Var framework is silent on this question. Consequently, we now introduce a new measure for the effectiveness of an observation.

Definition 4.1: An observation $\mathbf{z}(k) \in R^m$ at time k is said to be effective if the magnitude of the orthogonal projection $\|\hat{\mathbf{g}}\|$ is large. If $\|\hat{\mathbf{g}}\|$ is small, then the observation is said to be ineffective.

It turns out that $\|\hat{\mathbf{g}}\|$ can be small due to two mutually exclusive reasons. First, taking the norm of both sides of (4.2) and using (4.3), it is immediate that

$$\|\hat{\mathbf{g}}\| \leq \|\mathbf{g}\| \leq \|\mathbf{G}\| \|\mathbf{f}\|. \tag{4.4}$$

Hence, if the magnitude of the elements $G_{i,j}$, of the matrix \mathbf{G} are simultaneously small, then $\|\mathbf{G}\|$ is small and hence $\|\hat{\mathbf{g}}\|$ is small.

Stated in other words, if we place the observations in locations where the model is least sensitive, then from (4.4) it follows that the norm $\|\mathbf{g}\|$ is small in and around these locations the cost functional has flat patches in the control space R^{n+p} . Consequently, if $\|\hat{\mathbf{g}}\|$ is small, then the gradient method will bring very little improvement to forecast quality.

There is yet another situation where $\|\hat{\mathbf{g}}\|$ can be very small. This happens when the control error vector \mathbf{f} lies either in the null space⁴ of \mathbf{G} or very close to it. In such cases, \mathbf{g} is either orthogonal or nearly orthogonal to \mathbf{f} and so $\hat{\mathbf{g}}$ is very small even if $\|\mathbf{g}\|$ is large.

It is against this backdrop, we describe a framework in appendix A that directly relates the spectral properties of the matrix \mathbf{G} and its attendant impact on $\hat{\mathbf{g}}$.

Since \mathbf{f} is unknown but a constant vector, from (A.19) we get a basic relation:

⁴Let $\mathbf{A} \in R^{n \times n}$, then the null space $\text{NULL}(\mathbf{A})$ of the matrix \mathbf{A} is a subspace of R^n consisting of all vectors $\mathbf{x} \in R^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$.

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} = \sum_i d_i y_i^2 \geq \min_{y_i^2 \neq 0} \{y_i^2\} \left(\sum_i d_i \right). \tag{4.5}$$

From (A.20), we can express

$$\sum_i d_i = \text{tr}[\mathbf{G}] = \text{tr}[\overline{\mathbf{F}}^T \overline{\mathbf{F}}] = \|\overline{\mathbf{F}}\|_{\mathbf{F}}^2, \tag{4.6}$$

which is the Frobenius norm of $\overline{\mathbf{F}}$.

Combining (4.5) and (4.6), it is now clear that we can maximize $\|\hat{\mathbf{g}}\|$ by maximizing the Frobenius norm of the sensitivity matrix $\overline{\mathbf{F}}$ and this is independent of \mathbf{f} .

In the following section 5, we systematically explore the conditions for maximizing $\|\hat{\mathbf{g}}\|$ using (4.5) and (4.6).

5. Impact of observation placement

In exploring the properties of the projection $\hat{\mathbf{g}}$ and consequently the shape of the cost functional $\mathbf{J}(\mathbf{c})$ through dependence of its gradient on the forward sensitivities, we now move to analysis of expressions in (3.12)–(3.16).

Case 1: $\mathbf{x}(0)$ is varied and α is held constant.

In this case, $\delta\alpha = \mathbf{0}$ and from (3.14)–(3.16) and appendix B,

$$\mathbf{g} = -\nabla_{\mathbf{x}(0)} J = [\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U}] \delta \mathbf{x}(0) = [\overline{\mathbf{U}}^T \overline{\mathbf{U}}] \delta \mathbf{x}(0). \tag{5.1}$$

Case a: $m \geq n$ and $\text{Rank}(\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U}) = n$ (Meyer 2000). Then $\text{NULL}(\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U}) = \{\mathbf{0}\}$.

From (4.5) and (4.6), it follows that

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} \geq \min_{y_i^2 \neq 0} \{y_i^2\} \sum_j \|\overline{\mathbf{U}}_{*j}\|^2 \|\overline{\mathbf{U}}_{*j}\|^2, \tag{5.2}$$

where $\overline{\mathbf{U}}_{*j}$ is the j th column of $\overline{\mathbf{U}}$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in R^n$ are coordinates of $\delta \mathbf{x}(0)$ in the orthogonal basis formed by the eigenvectors of the Gramian $\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U}$. In (5.2), while we do not have any control over the magnitude of y_i^2 's, we can maximize the first factor on the right-hand side of (5.2) by placing the observations where the sum of the squares of the norms of the columns of $\overline{\mathbf{U}}$ is a maximum.

Case b: $m < n$ and $\text{rank}(\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U}) = m$. Then $\text{NULL}(\mathbf{U}^T \overline{\mathbf{H}} \mathbf{U})$ has dimension $(n \times m)$. For any $\delta \mathbf{c}$ in this null space, $\mathbf{g} \equiv \mathbf{0}$.

Again from (4.5) and (4.6), it is immediate that

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} \geq \min_{y_i^2 \neq 0} \{y_i^2\} \sum_j \|\mathbf{U}_{*j}\|^2. \tag{5.3}$$

The second factor on the right-hand side of (5.3) can be maximized by placing the observations where the

sum of the squares of the norms of the first m columns of \mathbf{U} is a maximum.

Case 2: α is varied and $\mathbf{x}(0)$ is held constant.

In this case $\delta\mathbf{x}(0) = 0$ and from (3.14)–(3.16), $\mathbf{g} = -\nabla_{\alpha} J = [\mathbf{V}^T \mathbf{H} \mathbf{V}] \delta\alpha$ where we recall from appendix B that $\mathbf{V}^T \mathbf{H} \mathbf{V} = \nabla^T \nabla \in R^{p \times p}$. Again, two cases arise depending on rank $(\mathbf{V}^T \mathbf{H} \mathbf{V}) = p$ or rank $(\mathbf{V}^T \mathbf{H} \mathbf{V}) < p$. In the latter case, the null space of $\mathbf{V}^T \mathbf{H} \mathbf{V}$ has dimension greater than or equal to one and for any $\delta\mathbf{c}$ in this null space, $\mathbf{g} \equiv 0$.

From (4.5) and (4.6), it can be verified that

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} \geq \min_{y_i^2 \neq 0} \{y_i^2\} \sum_j \|\bar{\mathbf{V}}_{*j}\|^2. \tag{5.4}$$

By placing the observations in the time interval where the sum of the squares of the norms of the matrix is a maximum, we can enforce or guarantee a nonzero lower bound on the norm of the adjoint gradient.

Case 3: Both $\mathbf{x}(0)$ and α are varied.

In this case, from (3.14)–(3.16), we get

$$\mathbf{g} = -\nabla_{\mathbf{c}} J = (\mathbf{F}^T \mathbf{H} \mathbf{F}) = (\mathbf{F}^T \mathbf{F}) \delta\mathbf{c}. \tag{5.5}$$

From (4.5) and (4.6), we get

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} \geq \min_{y_i^2 \neq 0} \{y_i^2\} \left(\sum_i d_i \right). \tag{5.6}$$

Hence, by placing observations in the time intervals containing the maximum of the sum of the square of the norms of the columns of \mathbf{F}^T , we bound the norm of $\hat{\mathbf{g}}$ as far away from zero as is feasible.

As a guide to applying the theory developed above, we summarize the key steps as follows.

Algorithm for the placement of observations

- Step 1: Pick a control $\mathbf{c} = [\mathbf{x}^T(0), \alpha^T]^T$ and run the model (2.1) forward in time to generate the forecast trajectory $\mathbf{x}(k)$ for $k \in [0, T]$ for some large $T > 0$.
- Step 2: Compute the Jacobians: $\mathbf{D}_{\mathbf{M}}(k)$, $\mathbf{D}_{\mathbf{M}}^{\alpha}(k)$ and $\mathbf{D}_{\mathbf{h}}(k)$ in (2.3) along the trajectory in step 1.
- Step 3: Compute the evolution of the forward sensitivity fields $\mathbf{U}(k)$ and $\mathbf{V}(k)$ using linear, time-varying dynamics in (3.2) and (3.3).
- Step 4: Compute $\mathbf{G} = \mathbf{F}^T \mathbf{H} \mathbf{F}$ using (3.15) and (3.16).
- Step 5: Plot the variation of one of the following norms versus time: (i) $\|\mathbf{G}\|_{\mathbf{F}}^2$, (ii) $\|\mathbf{G}_{*j}\|^2$, or (iii) G_{ij}^2 .
- Step 6: Identify small-time intervals containing the maxima in the plots generated in step 5. By sorting the values of these maxima in the decreasing order,

TABLE 1. Sensitivity of the solution of Eq. (6.1).

Sensitivities	Expression
$U = \partial x(t)/\partial x(0)$	$e^{-\beta t}$
$V_1 = \partial x(t)/\partial x_s$	$1 - e^{-\beta t}$
$V_2 = \partial x(t)/\partial \beta$	$\eta t e^{-\beta t}$
\mathbf{F}	$[U, V_1, V_2] \in R^{1 \times 3}$
$\mathbf{F}^T \mathbf{H} \mathbf{F}$	$(1/\sigma^2) \begin{bmatrix} U^2 & UV_1 & UV_2 \\ UV_1 & V_1^2 & V_1 V_2 \\ UV_2 & V_1 V_2 & V_2^2 \end{bmatrix}$
$\partial J/\partial x(0) = -g_1$	$-(1/\sigma^2)[U^2 \delta x(0) + UV_1 \delta x_s + UV_2 \delta \beta]$
$\partial J/\partial x_s = -g_2$	$-(1/\sigma^2)[UV_1 \delta x(0) + V_1^2 \delta x_s + V_1 V_2 \delta \beta]$
$\partial J/\partial \beta = -g_3$	$-(1/\sigma^2)[UV_2 \delta x(0) + V_1 V_2 \delta x_s + V_2^2 \delta \beta]$

we can then determine the number and location of the observations.

Clearly, those time intervals containing the maximum of one or more of the quantities in step 5 of the algorithm given above are the preferred locations for the placement of observations. The idea is that, if for any reason if we cannot place an observation at the maximum, we have the additional leverage of placing it close to it within a small interval containing it.

6. Two examples

In this section, we illustrate the theory developed in the previous sections using two examples. First is the 1D problem of air–sea interaction (Lakshminarayanan and Lewis 2010) where the control consists of initial and boundary conditions and a parameter. The second is the now famous example of the Saltzman’s 7D lower-order model for convection called SLOM (7).

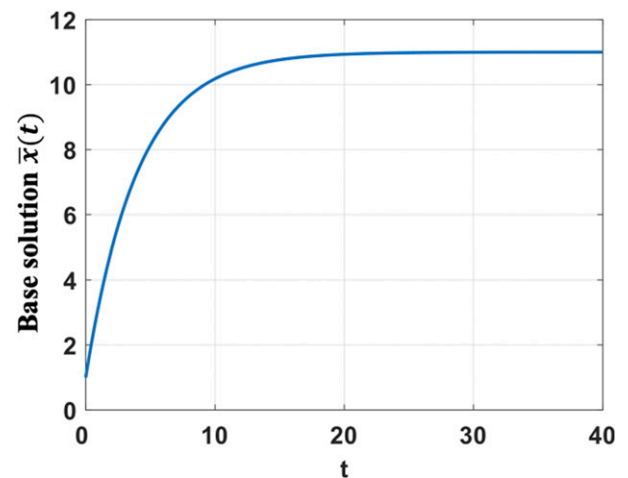


FIG. 3. The base solution $\bar{\mathbf{x}}(t)$ of (7.1) starting from the base control $\bar{\mathbf{c}} = [\bar{\mathbf{x}}(0), \bar{\alpha}, \bar{\beta}]^T = (1.0, 11, 0.25)^T$.

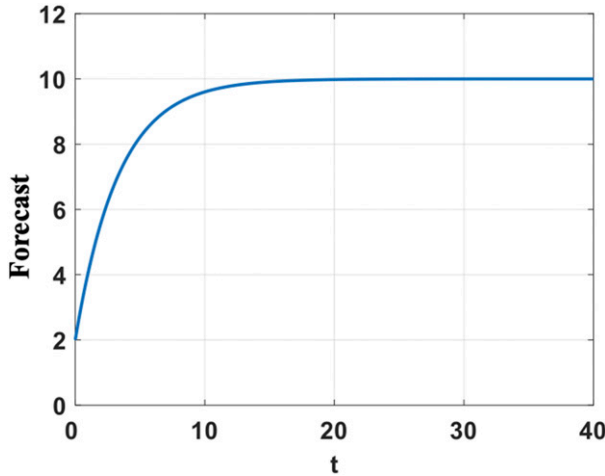


FIG. 4. Forecast $\mathbf{x}(t)$ using (7.1) starting from the control $\mathbf{c} = [x(0), x_s, \beta]^T = (2.0, 10, 0.3)^T$.

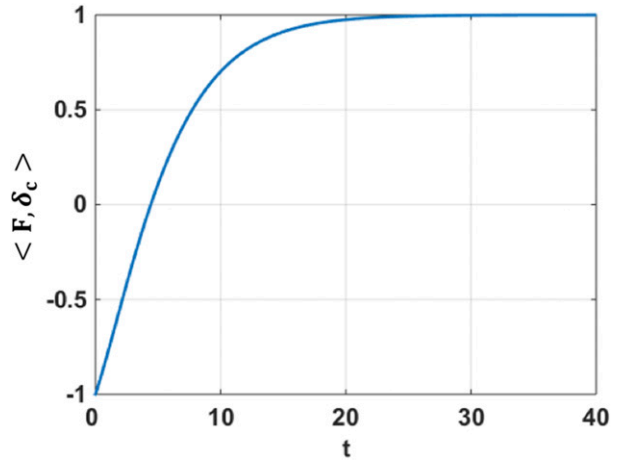


FIG. 6. Evolution of the inner product of \mathbf{F} in Table 1 and a fixed $\delta \mathbf{c} = (-1.0, 1.0, -0.05)^T$.

It is not out of place to mention that it is the initial analysis of SLOM (7) in Saltzman (1962) provided the key idea and the impetus to the discovery of the now famous Lorenz LOM (3) (Lorenz 1963) exhibiting deterministic chaos (Lorenz 1993, p. 137). Refer to the recent paper by Lakshmivarahan et al. (2019a) for more details on the relation between Saltzman (1962) and Lorenz (1963).

a. Air-sea interaction problem: 1D model

Let $n = 1$ and $p = 2$. Let $\alpha = (x_s, \beta)^T \in R^2$ be the set of two parameters. Consider the 1D model equation:

$$\frac{dx(t)}{dt} = \beta[x_s - x(t)], \tag{6.1}$$

with $x(0)$ as the initial condition. The control vector $\mathbf{c} = [x(0), \alpha^T]^T \in R^3$.

This model describes the dynamics of evolution of the normalized temperature $\mathbf{x}(t)$ of the cold continental

air as it moves with the prevailing wind over the warm ocean that is assumed to have a constant sea surface temperature x_s . Clearly, x_s denotes the boundary condition and β is known as the turbulent heat exchange parameter. See Lakshmivarahan and Lewis (2010) for more details.

The solution of (6.1) is given by

$$x(t) = x[t, x(0), x_s, \beta] = [x(0) - x_s]e^{-\beta t} + x_s = x_s - \eta e^{-\beta t}, \tag{6.2}$$

where $\eta = x_s - x(0)$. Exact expression for the forward sensitivities U, V_1, V_2 , the matrix $\mathbf{F}^T \mathbf{H} \mathbf{F}$, and expressions for the components of the adjoint sensitivities are given in Table 1.

A plot of the base solution $\bar{x}(t)$ of (6.1) starting from the base control $\bar{\mathbf{c}} = [\bar{x}(0), \bar{x}_s, \beta] = (1.0, 11, 0.25)^T$ is given in Fig. 3. In this and the rest of the figures, the discrete time k corresponds to the continuous time $t = k\Delta t$ where $\Delta t = 0.1$. Let $\mathbf{c} = [x(0), x_s, \beta]^T = (2.0, 10, 0.3)^T$

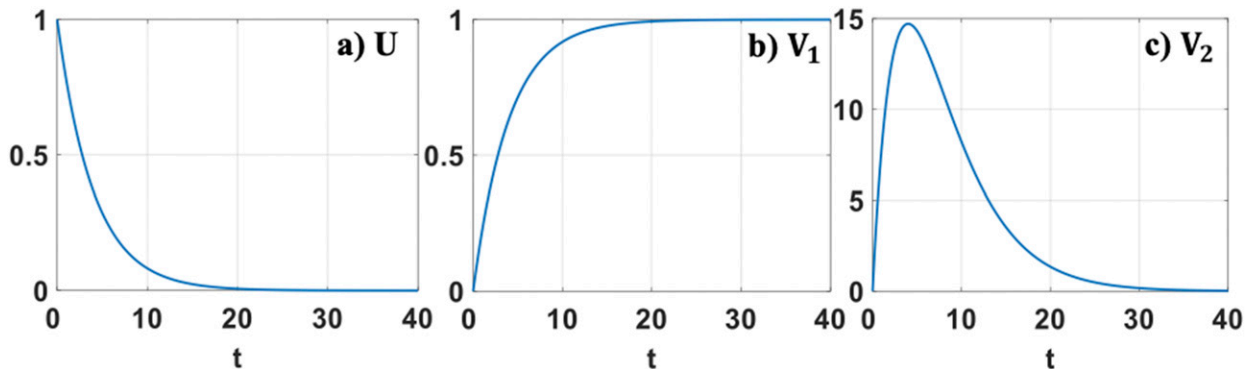


FIG. 5. Evolution of the forward sensitivities (a) U , (b) V_1 , and (c) V_2 given in Table 1 starting from the base control $\bar{\mathbf{c}}$ as in Fig. 4.

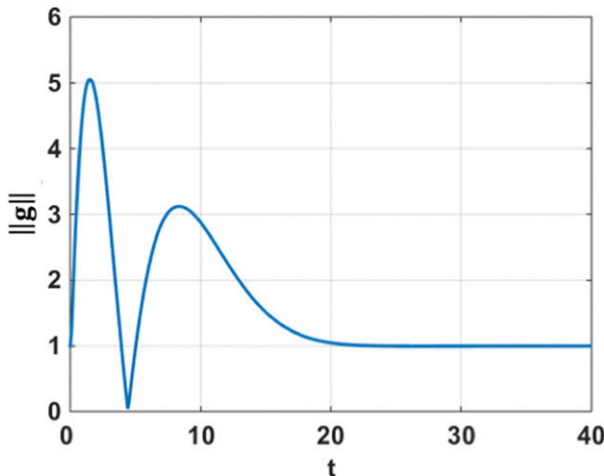


FIG. 7. Variation of the norm of adjoint gradient in Table 1 with time of observation.

be control for the forecast and the model forecast $x(t)$ using (6.1) starting from \mathbf{c} is plotted in Fig. 4. Evolution of forecast sensitivities of solution of (6.1) for the same initial condition as in Fig. 4 are given in Fig. 5.

Variation of the inner product of the time varying (row) sensitivity vector \mathbf{F} in Table 1 and the constant control error vector $\mathbf{f} = \delta\mathbf{c} = \bar{\mathbf{c}} - \mathbf{c}$, with time is given in Fig. 6. It can be verified from Fig. 6 that at time close to $t^* \approx 4.4$, the sensitivity vector \mathbf{F} is orthogonal to $\delta\mathbf{c}$ as indicated by the vanishing of the inner product. That is, $\delta\mathbf{c}$ at this time instant lies in the null space of the rank-one, symmetric matrix, $\mathbf{G} = \mathbf{F}^T \mathbf{H} \mathbf{F}$ given in Table 1. The plot of the variation of the norm of the gradient, $\|\mathbf{g}\|$ that is given in Table 1 is again illustrated in Fig. 7. Clearly, $\|\mathbf{g}\|$ vanishes at $t^* \approx 4.4$ and is small in its neighborhood.

Variation of the square of the individual forward sensitivities U , V_1 , and V_2 in Table 1 with time are given in Fig. 8. Clearly U^2 is a maximum close to the initial

time and V_1^2 attains its maximum value when t is large. But V_2^2 attains its maximum at time $t^{**} \approx (1/k) = 3.33$ (since $\beta = 0.3$ is used for forecast). Hence by placing three observations at times, say at times $t_1 = 1$, $t_2 = 20$, and $t_3 = 3.3$, we can capture information on the initial condition $x(0)$, boundary condition x_s and the parameter β . Combining (4.6) with expressions in Table 1, it can be verified (assuming $\sigma^2 = 1$) that $\text{tr}(\mathbf{G}) = \text{tr}(\mathbf{F}^T \mathbf{H} \mathbf{F}) = U^2 + V_1^2 + V_2^2$. Since U , V_1 , and V_2 vary in their magnitude, to bring out the key features in the sum, a plot of the scaled version: $(1/50)(U^2 + V_1^2 + V_2^2)$ against time is contained in Fig. 9. One can pick an equivalent set of three-time instances using this plot in Fig. 9 for placing observations.

We conclude this discussion by reporting the results of two experiments in Table 2. In the first experiment, four observations are chosen at times where the sensitivities are large and in the second where the sensitivities are small. The entries in the last column of Table 1 clearly illustrate the theory developed in this paper.

b. Application to Saltzman's Model: SLOM (7)

The strategy for numerical experiments follows the “identical twin” approach where two sets of model controls are specified—in this case of Saltzman’s spectral model, the Rayleigh number (Rayleigh 1916) and the model’s initial condition. One set is labeled “true control” and the other set, purposely made different than those for the true state is labeled “forecast control.” Variational data assimilation is used to correct the forecast control and bring it closer to true control. The theme of the experiment is to choose observation points that guarantee avoidance of “flatness” of the cost function in the space of control. The choice of these points rests on the foregoing theory that identifies a matrix and its trace. The location of these points is based on a two-part process:

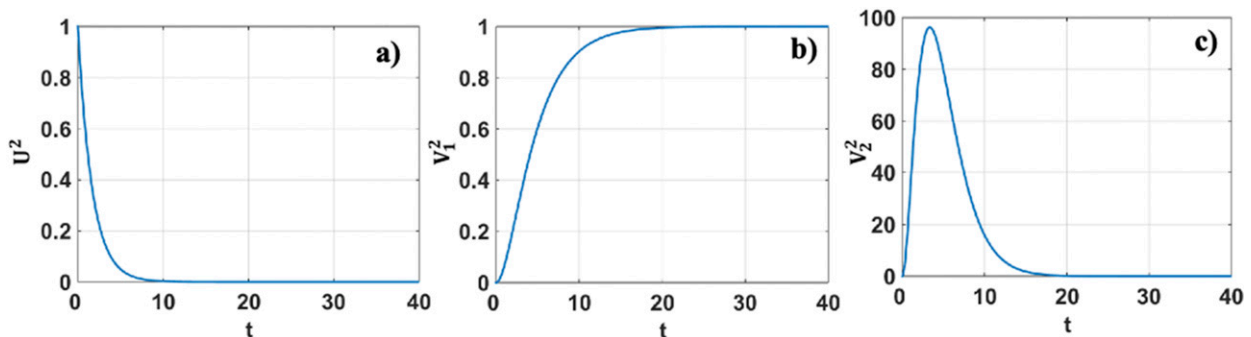


FIG. 8. Time evolution of U^2 , V_1^2 , and V_2^2 given in Table 1.

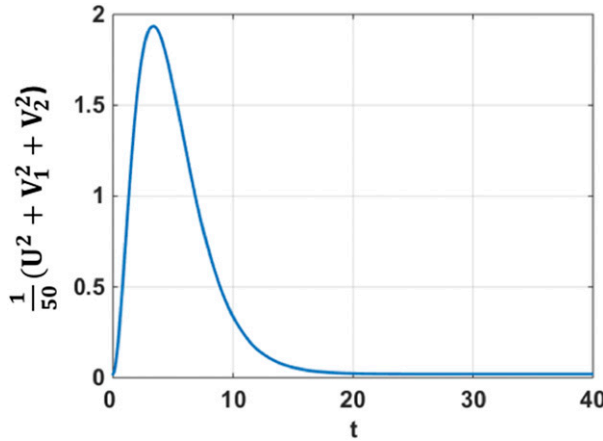


FIG. 9. Evolution of the sum of U^2 , V_1^2 , and V_2^2 given in Table 1.

1) PART 1: CHOICE OF LOCATION

(i) Stage 1: Controls specified

The true control $\bar{\mathbf{c}}$ and forecast control \mathbf{c} are specified. The evolution of the true state and forecast state are followed for a sufficiently long time to note differences in the convective regimes (as in step 1 of the algorithm in section 5).

(ii) Stage 2: Determination of forecast sensitivities and associated \mathbf{G} matrix

Forecast sensitivities to control, \mathbf{c} are found through solution to the augmented set of governing equations for Saltzman’s model. From these sensitivities and the forward operator [see section 6b(4) below]—the operator that determines the model counterpart to observations—the \mathbf{G} matrix and $\text{tr}(\mathbf{G})$, are determined. In the computation of \mathbf{G} and its trace, only the forward sensitivities to control and model counterpart of the observations are used, actual observations are not needed.

(iii) Stage 3: Observation placement

Observation locations are chosen in the (x, z, t) space where centroids of large values of $\text{tr}(\mathbf{G})$ are the ideal locations.

2) PART 2: SUBSEQUENT DATA ASSIMILATION

(i) Stage 4: Cost function J

The cost function J is defined as the sum of squared differences between the observations and the associated forecasts at the points in time and space where observation locations are chosen (observations at that point in time are based on true control and forecasts are based on erroneous control).

TABLE 2. Two experiments about the air–sea interaction problem.

No. of observations	Time of observation	$\ \delta\mathbf{c}\ $	$\mathbf{g}, \hat{\mathbf{g}}$, and angle θ between them
4	1, 2, 9, 10 (large sensitivities)	1.4151	$g = -3.689$ $\hat{g} = -1.254$ $\theta = 70^\circ$
4	4.5, 4.6, 4.7, 4.8 (small sensitivities)	1.4151	$g = -0.1849$ $\hat{g} = -0.0014$ $\theta = 89.57^\circ$

(ii) Stage 5: Variational data assimilation

The variational data assimilation method is used to adjust control and bring forecast control closer to true control.

Remark 6.1: Knowledge of the sensitivities found in step 2 allows calculation of cost-functional gradient without adjoint equations—an alternate method of 4D-Var. This alternate method is used in the numerical experiments that follow.

3) GOVERNING EQUATIONS

Saltzman (1962) assumed a seven-mode spectral model of convection that obeys the Boussinesq (1903) equations describing “roll” convection between two free surfaces maintained at constant temperature difference. Convection takes place in a thin layer of fluid heated from below. The convective motions take place in the x – z plane; that is, there is no variation in the y direction where x and y represent the horizontal Cartesian axes. The vertical coordinate z is perpendicular to the horizontal plane (positive upward). The momentum, mass, and thermodynamic equations can be reduced to two equations in vorticity and temperature departure—departure from the linear base-state temperature profile. These equations are found in Saltzman [1962, Eqs. (16) and (17)].

A double Fourier series form of spectral solution is assumed with three streamfunction modes and four temperature departure modes that are given in Table 3. Saltzman used an alphabetic representation for the seven spectral amplitudes (A, B, \dots, G), (section 7 in

TABLE 3. Seven spectral components of double Fourier series.

Variables	2D spectral components
Stream function ψ : Three components	$\psi = \begin{bmatrix} -4X_1 \sin(\pi ax) \sin(\pi z) \\ -4X_2 \sin(\pi bx) \sin(\pi z) \\ -4X_3 \sin(\pi cx) \sin(2\pi z) \end{bmatrix}$
Temperature θ : Four components	$\theta = \begin{bmatrix} 4X_4 \cos(\pi ax) \sin(\pi z) \\ 4X_5 \cos(\pi bx) \sin(\pi z) \\ 4X_6 \cos(\pi cx) \sin(2\pi z) \\ 4X_7 \sin(2\pi z) \end{bmatrix}$

TABLE 4. Boussinesq equations. These governing equations neglect variations in density except where they modify the action of gravity (Malkus and Veronis 1958; Chandrasekhar 1961, chapter 2).

Equations	Explanation of terms
$\partial \nabla^2 \psi / \partial t = -\partial(\psi, \nabla^2 \psi) / \partial(x, z) + \nu \nabla^2 \psi + g\alpha(\partial \theta / \partial x)$ $\partial \theta / \partial x = -\partial(\psi, \theta) / \partial(x, z) + (\Delta T / H)(\partial \psi / \partial x) + \kappa \nabla^2 \theta$	<p>In the above equations ψ is the two-dimensional streamfunction and θ the departure of temperature from base state. The fluid velocity components u (x direction) and w (z direction) are given by $(u, w) = [-(\partial \psi / \partial z), \partial \psi / \partial x]$, vorticity is given by $\nabla^2 \psi = \partial u / \partial z - \partial w / \partial x$, $\nabla^4 \psi = \nabla^2(\nabla^2 \psi) = \partial^4 \psi / \partial x^4 + \partial^4 \psi / \partial z^4 + 2(\partial^4 \psi / \partial x^2 \partial z^2)$, and the Jacobian $\partial(A, B) / \partial(x, z) = (\partial A / \partial x)(\partial B / \partial z) - (\partial A / \partial z)(\partial B / \partial x)$ accounts for the fluid advection of vorticity and temperature when $A = \psi$ and $B = \nabla^2 \psi$ or θ.</p>

Saltzman 1962), whereas we adopted the representation X_1 to X_7 for the amplitudes.

The following trigonometric arguments appear in the spectral forms: πax , πbx , and πcx with $a = 1/\sqrt{2}$, $b = 2\sqrt{2}/3$, and $c = \sqrt{2}/6$ where they respectively account for three waves, four waves, and one wave over the nondimensional $6\sqrt{2}$ horizontal length of the domain, and the arguments πz and $2\pi z$ represent a half wave and a full wave over the depth of fluid where nondimensional depth to length ratio is $1/6\sqrt{2}$.

Substitution of the spectral forms of solution into the Boussinesq equations in Table 4 leads to a set of coupled ordinary differential equations governing the spectral amplitudes. These equations and associated coefficients are found in Tables 5 and 6. The coefficients in Table 6 are expressed in terms of the two nondimensional numbers associated with this convection problem: Rayleigh number (λ) and Prandtl number (σ). The Rayleigh number is expressed as a constant λ multiplying the critical Rayleigh number in Saltzman's development [=657.511 for this case of two free boundaries as derived in Chandrasekhar (1961) that followed Rayleigh's pioneering development in 1916]]. Saltzman set $\sigma = 10$ in his study and we also set $\sigma = 10$ in our study, but only after we determined that the forecast sensitivity to σ was much less than sensitivity to λ . Determination of the relative

forecast sensitivities to these parameters dictated our derivation in terms of arbitrary values of the two numbers.

In this study, the true control vector is taken to be $\lambda = 2$ and Saltzman's specified initial conditions. The forecast control vector is taken to be $\lambda = 2.1$ and Saltzman's specified initial conditions except for temperature-departure amplitude $X_4(0)$ that is taken to be nondimensional number 0.1 instead of 0.0. It is sufficient to choose the single amplitude initial condition, different from true control, to reveal the action of the variational data assimilation process on correction of initial condition.

TABLE 5. Saltzman's (1962) model: SLOM (7) is $\dot{\mathbf{X}} = \mathbf{f}(\mathbf{X}, \boldsymbol{\alpha})$, where $\mathbf{X} \in R^7$, $\boldsymbol{\alpha} = (\lambda, \sigma)^T$ and $\mathbf{f} = (f_1, f_2, \dots, f_7)^T$.

Components of \mathbf{f}	Their expressions
$f_1(\mathbf{X}, \boldsymbol{\alpha})$	$C_{123}X_2X_3 + C_{140}X_4 + C_{110}X_1$
$f_2(\mathbf{X}, \boldsymbol{\alpha})$	$C_{213}X_1X_3 + C_{250}X_5 + C_{220}X_2$
$f_3(\mathbf{X}, \boldsymbol{\alpha})$	$C_{321}X_1X_2 + C_{360}X_6 + C_{330}X_3$
$f_4(\mathbf{X}, \boldsymbol{\alpha})$	$C_{435}X_3X_5 + C_{426}X_2X_6 + C_{417}X_1X_7$ $+ C_{410}X_1 + C_{440}X_4$
$f_5(\mathbf{X}, \boldsymbol{\alpha})$	$C_{534}X_3X_4 + C_{516}X_1X_6 + C_{527}X_2X_7$ $+ C_{520}X_2 + C_{550}X_5$
$f_6(\mathbf{X}, \boldsymbol{\alpha})$	$C_{615}X_1X_5 + C_{624}X_2X_4 + C_{630}X_3 + C_{660}X_6$
$f_7(\mathbf{X}, \boldsymbol{\alpha})$	$C_{714}X_1X_4 + C_{725}X_2X_5 + C_{770}X_7$

TABLE 6. Values of coefficients C_{ijk} in Table 5.

$C_{123} = -\pi^2[(2b - c)(c^2 - b^2 + 3)/(a^2 + 1)] = 23.521$
$C_{140} = -\sigma[a/\pi(a^2 + 1)] = -0.15\sigma$
$C_{110} = -\sigma\pi^2(a^2 + 1) = -14.80\sigma$
$C_{213} = -\pi^2[(2a + c)(a^2 - c^2 - 3)/(b^2 + 1)] = -22.030$
$C_{250} = -\sigma[b/\pi(b^2 + 1)] = -0.159\sigma$
$C_{220} = -\sigma\pi^2(b^2 + 1) = -18.643\sigma$
$C_{312} = \pi^2[(b - a)(a + b)^2/(c^2 + 4)] = 1.561$
$C_{360} = -\sigma[c/\pi(c^2 + 4)] = -0.0185\sigma$
$C_{330} = -\sigma\pi^2(c^2 + 4) = -40.267\sigma$
$C_{435} = -\pi^2(2b - c) = -16.284$
$C_{426} = -\pi^2(2b - c) = -16.284$
$C_{417} = -4\pi^2a = -27.916$
$C_{440} = -\pi^2(a^2 + 1) = -14.804$
$C_{410} = -\pi a R_c \lambda = 1460.62\lambda$
$C_{534} = \pi^2(2a + c) = 16.284$
$C_{516} = -\pi^2(2a + c) = -16.284$
$C_{527} = -4\pi^2b = -37.221$
$C_{520} = -\pi b R_c \lambda = -1947.508\lambda$
$C_{550} = -\pi^2(b^2 + 1) = -18.643$
$C_{615} = \pi^2(a + b) = 16.284$
$C_{624} = \pi^2(a + b) = 16.284$
$C_{630} = -\pi c R_c \lambda = -486.877\lambda$
$C_{660} = -\pi^2(c^2 + 4) = -40.028$
$C_{714} = 2\pi^2a = 13.956$
$C_{725} = 2\pi^2b = 18.610$
$C_{770} = -4\pi^2 = -39.479$

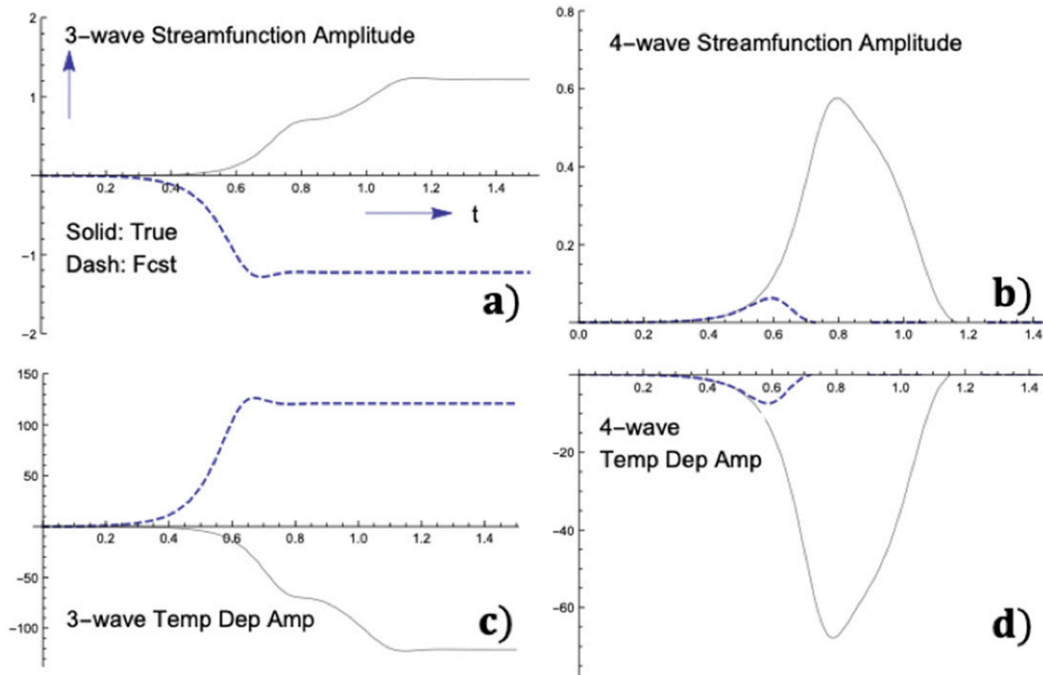


FIG. 10. Time evolution of true and forecasted spectral amplitudes.

In addition to the spectral equations for convection two sets of equations governing the forecast sensitivities to control are derived, namely, equations governing $\partial X_i / \partial \lambda$, $[\partial X_i / \partial X_4(0)]$ $i = 1, 2, \dots, 7$. Thus, there are 21 coupled equations where solutions to $X_i(t)$ equations are coupled with the time evolution of the 14 sensitivity equations but where the solutions to the sensitivity equations do not feed back into the solutions to the $X_i(t)$ equations. The graphs of amplitude variation for true and forecast control are shown in Fig. 10. Figures 11 and 12 show the (x, z) distribution of the true and forecasted vertical velocity and temperature departure variables (w, θ) at $t = 0.65$, the time when data assimilation operates. Note that the phase for wavenumber 3 is reversed for true and forecasted states in Figs. 11 and 12, consistent with the time evolution of amplitudes shown in Fig. 10.

4) THE EXPERIMENT

The forward operator \mathbf{H} , a matrix of dimension 3×7 for this model, represents the spatial form of (u, w, θ) variables in the (x, z) space (refer to Table 7 for details). When multiplied by the 7×1 time-dependent amplitude matrix $\mathbf{X} = \{X_1, X_2, \dots, X_7\}$ of Fourier amplitudes, the three rows of the resulting matrix represent the (x, z, t) structure of the (u, w, θ) model variables—the state of convection in space and time. When the \mathbf{H} matrix is multiplied by

the \mathbf{F} matrix, a 7×2 matrix of forecast sensitivities (the two columns are amplitudes of the forecast sensitivity to the initial condition $X_4(0)$ and Rayleigh number (expressed in terms of λ), respectively, the result is a 3×2 matrix of the (x, z, t) representation of the six sensitivity elements—the separate first derivatives of the variables with respect to the two control parameters. In matrix form, $\mathbf{H}\mathbf{X}$ is the state vector and $\mathbf{H}\mathbf{F}$ is the sensitivity matrix. The square of the sensitivity matrix, $(\mathbf{H}\mathbf{F})^T(\mathbf{H}\mathbf{F})$, is the \mathbf{G} matrix, and $\text{tr}(\mathbf{G})$ is its trace.

The variational data assimilation process will take place at $t = 0.65$, that point in time where initial perturbations in the Fourier convective components have started to grow significantly. The cost function J over the (x, z) space at this time is also calculated. The trace $\text{tr}(\mathbf{G})$ is plotted in Fig. 13.⁵ A point at the maximum value of the $\text{tr}(\mathbf{G})$ centroid near $x = 6.0$ and $z = 0.7$ will be considered as the location of observations of (u, w, θ) —the maximum is located at $x = 5.802$ and $z = 0.721$.

⁵ The graphic of the Frobenius norm has structure similar to trace \mathbf{G} but much greater magnitude. The relative difference in magnitudes is not a factor in identifying locations where sensitivities are large.

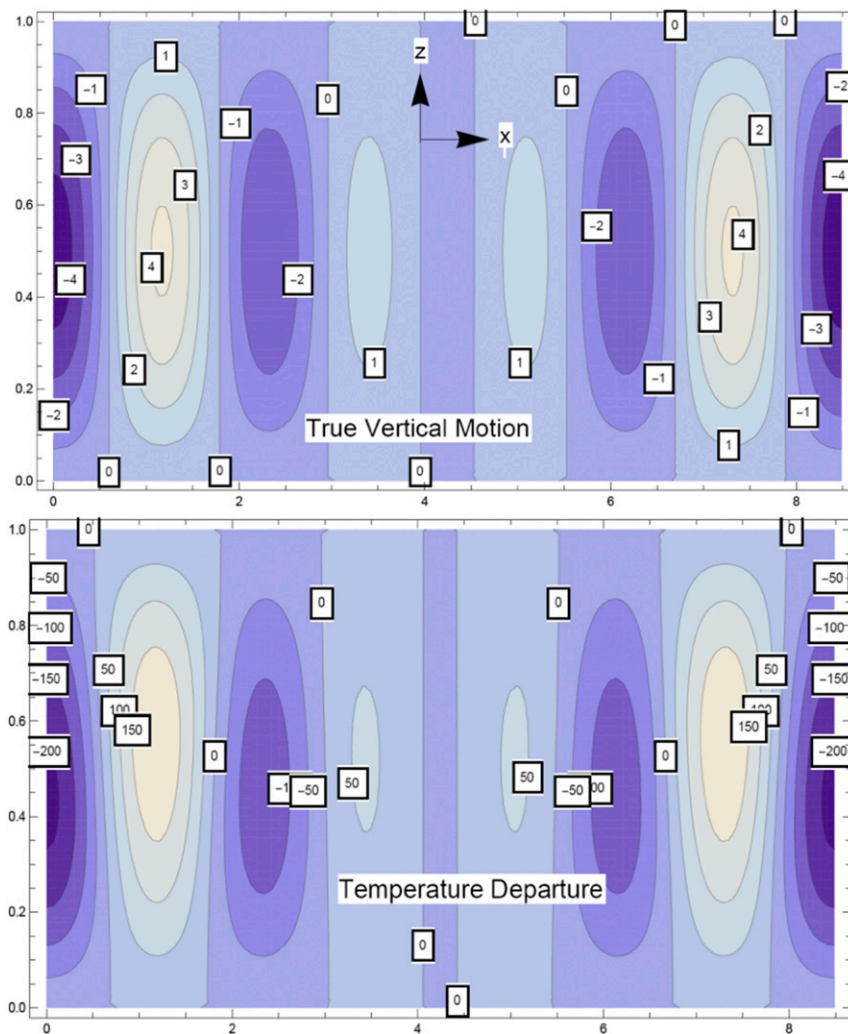


FIG. 11. The (x, z) distribution of (top) true vertical motion and (bottom) temperature departure at $t = 0.65$.

The two-component gradient of J , gradient of J with respect to λ and the gradient of J with respect to the initial condition $X_4(0)$, is found by taking the derivative of the cost function with respect to these controls separately. These derivatives can be calculated easily since the variables' sensitivity to control are available from solution to the amplitude equations discussed earlier. The structure of these two gradients over the (x, z) space at $t = 0.65$ is shown in Figs. 14 and 15.

The values of the gradients at the observation point are $\partial J / [\partial X_4(0)] = 4.89399 \times 10^5$ and $\partial J / \partial \lambda = 4.17800 \times 10^5$. Thus, from the operating point in the space of control $[X_4(0), \lambda] = (0.1, 2.1)$, The negative gradient of J is directed downward at an angle of 49.51° (slope of 1.17). With a step size of 0.150 along this direction, the

adjusted control is $[\Delta X_4(0), \Delta \lambda] = (-0.097, -0.110)$, close to the optimal adjustment where only a 5° difference exists between the two vectors. The adjusted forecast control $[\Delta X_4(0), \lambda] = (0.003, 1.99)$, is very close to true control.

In summary, the $\text{tr}(\mathbf{G})$ structure identified points in (x, z) space at $t = 0.65$ that were likely to have large values of the J gradient. When observations of the three convection variables (two velocity components and the temperature departure component) are placed at one of the centroids of the \mathbf{G} trace map, the gradients of the cost function were found to be large and of roughly the same magnitude. The direction of the negative gradient was very close to the optimal direction. Examination of the gradient maps in Figs. 14 and 15 makes it clear that there are many locations for observation points [away

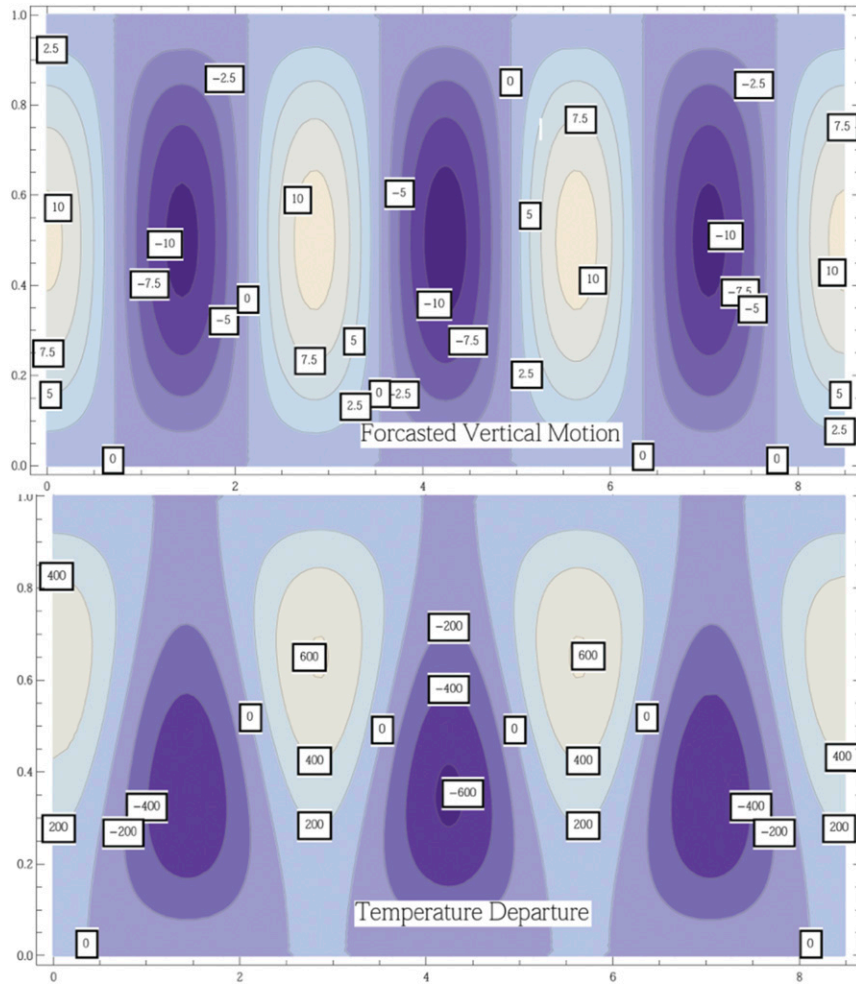


FIG. 12. The (x, z) distribution of (top) forecasted vertical motion and (bottom) temperature departure at $t = 0.65$.

from the centroids of $\text{tr}(\mathbf{G})$] that deliver gradients much smaller and in directions that are not pointed toward the J minimum compared to those associated with points near centroids.

7. Discussion and conclusions

Whereas traditional 4D-Var is grounded on use of the adjoint model to deliver the gradient of the cost function—so-called adjoint sensitivity, the alternate method developed here is founded on the role of forecast sensitivity to control and an expression that defines the forecast error in terms of observation error and the model counterpart to observations. The separability of the influence of the model and observation system as brought out in (4.1) on adjoint sensitivity presents a view that enriches the traditional view. We want to emphasize that since the error \mathbf{f} is not

known in advance, we use the fundamental relation in (4.1) to decide on the placement of the observation as described above.

The traditional 4D-Var method provides a computationally efficient framework for computing the adjoint gradient for use in forecast error correction. Within this classical framework, the prevailing practice is to use different types of observations, without much consideration to their optimal location. And as the number of observations increases, computational demands to assimilate them also increase. Our alternate approach presents an off-time strategy based on the forward sensitivity field in the control space. By selectively placing observations where the model exhibits large sensitivity, we can in fact increase the effectiveness of the observations without having to use many observations in the assimilation process.

TABLE 7. Elements of the matrix **H**.

Row of H	Matrix elements	Values
First row	$H(1, 1)$	$4\pi \sin(\pi ax) \cos(\pi z)$
	$H(1, 2)$	$4\pi \sin(\pi bx) \cos(\pi z)$
	$H(1, 3)$	$4\pi \sin(\pi cx) \cos(2\pi z)$
	$H(1, 4)$	0
	$H(1, 5)$	0
	$H(1, 6)$	0
	$H(1, 7)$	0
Second row	$H(2, 1)$	$-4\pi a \cos(\pi ax) \sin(\pi z)$
	$H(2, 2)$	$-4\pi b \cos(\pi bx) \sin(\pi z)$
	$H(2, 3)$	$-4\pi c \cos(\pi cx) \sin(2\pi z)$
	$H(2, 4)$	0
	$H(2, 5)$	0
	$H(2, 6)$	0
	$H(2, 7)$	0
Third row	$H(3, 1)$	0
	$H(3, 2)$	0
	$H(3, 3)$	0
	$H(3, 4)$	$4 \cos(\pi ax) \sin(\pi z)$
	$H(3, 5)$	$4 \cos(\pi bx) \sin(\pi z)$
	$H(3, 6)$	$4 \cos(\pi cx) \sin(2\pi z)$
	$H(3, 7)$	$4 \sin(2\pi x)$

Our approach also clearly shows that it is not simply the magnitude of the adjoint gradient, but its (orthogonal) projection along the unknown control error vector $\delta\mathbf{c}$ that is solely responsible for the reduction in the forecast error. It is shown that the magnitude of this projection is related to the inner product of the adjoint sensitivity and the unknown error in the control vector. This inner product is directly related to the Frobenius norm of the forward sensitivity of the forecast with respect to the initial conditions and parameters.

By examining variations of this norm in space and time, and by placing the observations in locations at or in the vicinity of the maxima of this norm, the magnitude of the projection is increased. The effectiveness of this approach has been demonstrated with a one-dimensional air/sea interaction model (time dependence alone) and a two-dimensional model, Saltzman’s nonlinear model of Bénard (1900, 1901) convection in space and time. The richness of Saltzman’s treatment of convection so faithful to the work of Bénard and Rayleigh was a pedagogical triumph, central to Ed Lorenz’s monumental study of deterministic chaos, and central to our understanding of the interplay between dynamics and observations in data assimilation.

The strategy of choosing individual observations in the vicinity of the Frobenius norm maxima is straightforward and logical, but a methodology that improves the “granularity” (finer resolution) can be attained by the following processes: 1) in the presence of an inordinate number of Frobenius norm maxima, rank order them in terms of the value of the maximum and decide on a suitable number of observations, 2) examine the square of the norm of each of the n columns of \mathbf{U} against time and space and identify the maxima in each field; then again by rank ordering these maxima, decide on a suitable number of observations to use, and 3) the finest level of granularity is attained when we examine the variation of the square of each of the $(n + p) \times m$ elements of \mathbf{U} , identify and rank order the maxima to come up with an acceptable number of observations to use in the assimilation scheme.

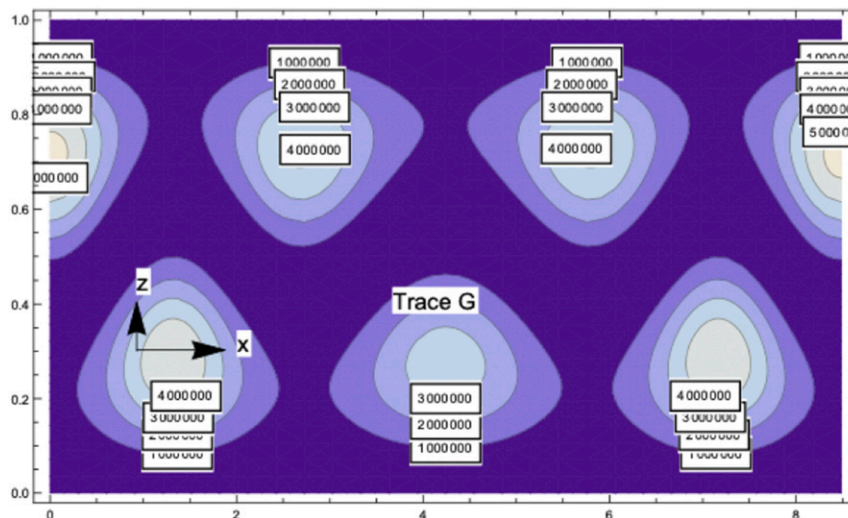


FIG. 13. Trace of **G** in the (x, z) space at $t = 0.65$.

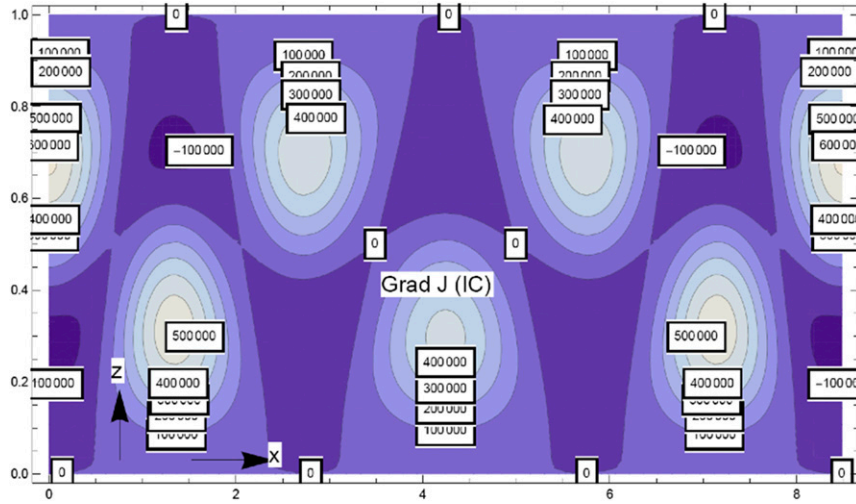


FIG. 14. The cost function gradient with respect to initial condition at $t = 0.65$.

We hasten to add that the process of computing the forward sensitivities involves solving time-consuming matrix recurrence equations in (3.2) and (3.3) where each step involves $O(n^3)$ operations (See Lakshmivarahan et al. 2017). There are two avenues to conquer this computational challenge. First is to invoke parallel implementation of the recurrence that defines the forward sensitivity dynamics to gain speedup. A little reflection suggests that Eqs. (3.2) and (3.3) admit “embarrassingly” parallel implementations. Once the Jacobian of the forward operators $\mathbf{D}_M(k)$ in (3.2) are computed along the forecast trajectory (which is a common denominator in all the adjoint based approaches), we can split the matrix–matrix recurrence in (3.2) into a system of

n matrix–vector recurrences split across the n columns of $\mathbf{U}(k)$ and implement each of these matrix–vector recurrences in parallel requiring $O(n^2)$ operations per iteration. Using this approach, computation time for this parallel implementation would indeed match that of running the tangent linear system. The second idea is to consider a multiscale, multigrid approach wherein we use a relatively small-sized coarse grid to compute the forward sensitivity matrices that are critical to decide on the location and the number of observations. Then use a finer-scale grid to assimilate the observation. Of course, we also have a choice to hybridize the multiscale, multigrid approach with parallel computing to make the overall process computationally feasible.

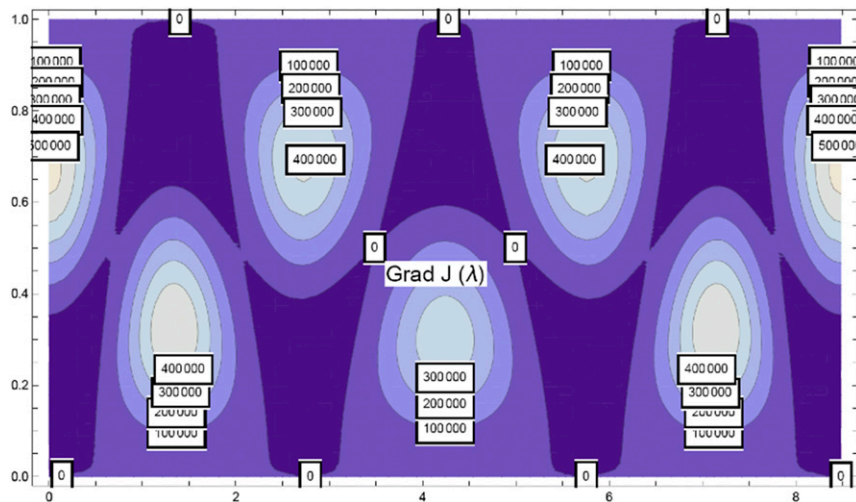


FIG. 15. The cost function gradient with respect to λ at $t = 0.65$.

Acknowledgments. This paper is a condensed version of a two-part paper (Lakshminarayanan et al. 2019b; Lewis et al. 2019) submitted to *J. Atmos. Sci.* The challenge of combining these two papers was made manageable through the advice and the input from the editor-in-charge and valued suggestions from a large panel of reviewers. We are indebted to these individuals for helping us create a better and a readable manuscript. Preliminary ideas related to the developments in this paper were presented (by S. L.) in a seminar in July 2018 at the Center for Atmospheric and Oceanic Sciences (CAOS) and Divecha Centre for Climate Change (DCCC) at the Indian Institute of Science. We are grateful to all the participants for the interaction and to Ashwin Seshadri, J. Srinivasan, and S. K. Satheesh for the hospitality and keen interest.

APPENDIX A

Dependence of $\|\hat{\mathbf{g}}\|$ on the Spectral Properties of $\mathbf{G} = \mathbf{F}^T \mathbf{H} \mathbf{F}$

In this appendix we explore the effectiveness of an observation by analyzing the intrinsic dependence of the projected component, $\hat{\mathbf{g}}$ of the adjoint gradient $\mathbf{g} = \mathbf{G}\mathbf{f}$ along the control error vector $\mathbf{f} = \delta\mathbf{c}$. We start by enlisting some of the key properties of the linear transformation defined by \mathbf{G} .

Property A1: Referring to the developments in section 3 and appendix B, it follows that matrix \mathbf{G} is a Gramian and is given by

$$\mathbf{G} = \mathbf{F}^T \mathbf{H} \mathbf{F} = \overline{\mathbf{F}}^T \overline{\mathbf{F}} \in R^{(n+p) \times (n+p)} \quad (\text{A.1})$$

and is an SPSD (Meyer 2000).

Property A2: Since \mathbf{G} is SPSD, the inner product of \mathbf{f} and \mathbf{g} is nonnegative. It follows from the definition of SPSD (Meyer 2000) that

$$\langle \mathbf{f}, \mathbf{g} \rangle = \langle \mathbf{f}, \mathbf{G}\mathbf{f} \rangle = \mathbf{f}^T \mathbf{G}\mathbf{f} = \|\mathbf{f}\|^2 (\hat{\mathbf{f}}^T \mathbf{G}\hat{\mathbf{f}}) \geq 0, \quad (\text{A.2})$$

where $\hat{\mathbf{f}} = \mathbf{f}/\|\mathbf{f}\|$ is the unit vector in the direction of \mathbf{f} .

Property A3: Referring to chapter 6 in Lewis et al. (2006), the orthogonal projection matrix $\mathbf{P}_{\mathbf{f}}$ along the span of the vector \mathbf{f} is given by

$$\mathbf{P}_{\mathbf{f}} = \mathbf{f}(\mathbf{f}^T \mathbf{f})^{-1} \mathbf{f}^T = \frac{\mathbf{f}\mathbf{f}^T}{\|\mathbf{f}\|^2} = \hat{\mathbf{f}}\hat{\mathbf{f}}^T. \quad (\text{A.3})$$

Property A4: The orthogonal projection $\hat{\mathbf{g}}$ of \mathbf{g} along \mathbf{f} is given by Lewis et al. (2006, chapter 6)

$$\hat{\mathbf{g}} = \mathbf{P}_{\mathbf{f}} \mathbf{g} = \hat{\mathbf{f}}\hat{\mathbf{f}}^T \mathbf{g} = \hat{\mathbf{f}}\hat{\mathbf{f}}^T (\mathbf{G}\mathbf{f}) = \mathbf{f}(\hat{\mathbf{f}}^T \mathbf{G}\hat{\mathbf{f}}). \quad (\text{A.4})$$

Property A5: Since \mathbf{f} and $\hat{\mathbf{f}}$ are fixed, the properties of the scalar

$$\Omega(\mathbf{G}) = \hat{\mathbf{f}}^T \mathbf{G}\hat{\mathbf{f}} \quad (\text{A.5})$$

called the Rayleigh coefficient (Meyer 2000) depends critically on the spectral expansion of the SPSD matrix \mathbf{G} .

Property A6: Since \mathbf{G} is SPSD, the spectral expansion of \mathbf{G} is given by (Meyer 2000)

$$\mathbf{G} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T, \quad (\text{A.6})$$

where \mathbf{Q} and \mathbf{D} are the matrices of eigenvectors and eigenvalues of \mathbf{G} , respectively, where $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{n+p}$. Substituting (A.6) in (A.5) and define

$$\mathbf{y} = \mathbf{Q}^T \hat{\mathbf{f}}, \quad (\text{A.7})$$

we get

$$\Omega(\mathbf{G}) = \hat{\mathbf{f}}^T \mathbf{G}\hat{\mathbf{f}} = \hat{\mathbf{f}}^T (\mathbf{Q}\mathbf{D}\mathbf{Q}^T) \hat{\mathbf{f}} = \mathbf{y}^T \mathbf{D}\mathbf{y} = \sum_i d_i y_i^2, \quad (\text{A.8})$$

where \mathbf{y} is a unit vector since

$$\|\mathbf{y}\|^2 = \mathbf{y}^T \mathbf{y} = \hat{\mathbf{f}}^T (\mathbf{Q}\mathbf{Q}^T) \hat{\mathbf{f}} = \hat{\mathbf{f}}^T \hat{\mathbf{f}} = \|\hat{\mathbf{f}}\|^2 = 1. \quad (\text{A.9})$$

Stated in other words, the components of \mathbf{y} are the coordinates of the unit vector $\hat{\mathbf{f}}$ in the new orthogonal coordinate system formed by the columns of \mathbf{Q} .

In view of (A.9), the Rayleigh coefficient $\Omega(\mathbf{G})$ has natural interpretation as the weighted average of the eigenvalues of \mathbf{G} .

Property A7: Using (A.2) in (A.4), we obtain another natural expression for the projection $\hat{\mathbf{g}}$ as

$$\hat{\mathbf{g}} = \mathbf{f} \frac{\langle \mathbf{f}, \mathbf{g} \rangle}{\|\mathbf{f}\|^2} = \hat{\mathbf{f}} \langle \hat{\mathbf{f}}, \mathbf{g} \rangle. \quad (\text{A.10})$$

A geometric view of the relation between \mathbf{f} , \mathbf{g} , and $\hat{\mathbf{g}}$ is given in Fig. 1. Referring to Fig. 1, we can resolve \mathbf{g} as a sum:

$$\mathbf{g} = \hat{\mathbf{g}} + \mathbf{g}^\perp, \quad (\text{A.11})$$

where $\hat{\mathbf{g}}$ is the orthogonal projection of \mathbf{g} onto \mathbf{f} and $\mathbf{g}^\perp = \mathbf{g} - \hat{\mathbf{g}}$.

The overall effectiveness of the forecast error correction strategy depends critically on $\hat{\mathbf{g}}$ through the length, $\|\hat{\mathbf{g}}\|$. Clearly, referring to Fig. 1, we can maximize $\|\hat{\mathbf{g}}\|$ by maximizing $\|\hat{\mathbf{g}}\|$ and minimizing angle θ between \mathbf{f} and \mathbf{g} that is equivalent to maximizing the inner product $\langle \mathbf{f}, \mathbf{g} \rangle$.

Property A8: Combining the eigen decomposition of \mathbf{G} in (A.6) with (A.2) and (A.7), we get the following new expression:

$$\mathbf{g} = \mathbf{G}\mathbf{f} = (\mathbf{Q}\mathbf{D}\mathbf{Q}^T)\mathbf{f} = \|\mathbf{f}\| (\mathbf{Q}\mathbf{D})\mathbf{Q}^T \hat{\mathbf{f}} = \|\mathbf{f}\| (\mathbf{Q}\mathbf{D})\mathbf{y}. \quad (\text{A.12})$$

Consequently,

$$\begin{aligned}\|\mathbf{g}\|^2 &= \mathbf{g}^T \mathbf{g} = \|\mathbf{f}\|^2 \mathbf{y}^T \mathbf{D} (\mathbf{Q}^T \mathbf{Q}) \mathbf{D} \mathbf{y} \\ &= \|\mathbf{f}\|^2 \mathbf{y}^T \mathbf{D}^2 \mathbf{y} = \|\mathbf{f}\|^2 \sum_i d_i^2 y_i^2,\end{aligned}\quad (\text{A.13})$$

where the second factor on the right-hand side of (A.13) is the weighted average of the squares of the eigenvalues of \mathbf{G} .

Property A9: Combining (A.4), (A.6), and (A.9), we get an expression for $\hat{\mathbf{g}}$:

$$\hat{\mathbf{g}} = \mathbf{f}(\hat{\mathbf{f}}^T \mathbf{G} \hat{\mathbf{f}}) = \mathbf{f}[\hat{\mathbf{f}}^T (\mathbf{Q} \mathbf{D} \mathbf{Q}^T) \hat{\mathbf{f}}] = \mathbf{f}(\mathbf{y}^T \mathbf{D} \mathbf{y}). \quad (\text{A.14})$$

Hence,

$$\|\hat{\mathbf{g}}\|^2 = \hat{\mathbf{g}}^T \hat{\mathbf{g}} = \|\mathbf{f}\|^2 (\mathbf{y}^T \mathbf{D} \mathbf{y})^2 = \|\mathbf{f}\|^2 \left(\sum_i d_i y_i^2 \right)^2, \quad (\text{A.15})$$

where the second factor on the right-hand side of (A.15) is the square of the weighted average of the eigenvalues of \mathbf{G} .

Property A10: Considering $\{y_i^2\}$ as the discrete probability distribution over d_i , define

$$\bar{\mathbf{d}} = \sum_i d_i y_i^2. \quad (\text{A.16})$$

Then, from

$$0 \leq \sum_i (d_i - \bar{\mathbf{d}})^2 y_i^2 = \sum_i d_i^2 y_i^2 - \bar{\mathbf{d}}^2, \quad (\text{A.17})$$

we obtain the natural inequality:

$$\|\mathbf{g}\|^2 \geq \|\hat{\mathbf{g}}\|^2, \quad (\text{A.18})$$

which states that the length of the vector \mathbf{g} is always greater than or equal to its orthogonal projection $\hat{\mathbf{g}}$.

The above analysis naturally leads to the following inescapable conclusions:

Property A11: Increasing $\|\hat{\mathbf{g}}\|$ has an immediate consequence of (i) increasing the effectiveness of the forecast error correction process and (ii) it increases $\|\mathbf{g}\|$, which in turn controls the shape of the cost functional $J(\mathbf{c})$ by avoiding flat patches. Hence, the reason for the title of the paper.

As a prelude to increasing $\|\hat{\mathbf{g}}\|$ from (A.15) first observe that for a given fixed \mathbf{f} ,

$$\frac{\|\hat{\mathbf{g}}\|}{\|\mathbf{f}\|} = \sum_i d_i y_i^2 \geq \min_{y_i^2 \neq 0} \{y_i^2\} \left(\sum_i d_i \right). \quad (\text{A.19})$$

From (A.6), exploiting the properties of the trace of product matrices (Meyer 2000), it can be verified that

$$\begin{aligned}\text{tr}(\mathbf{G}) &= \text{tr}(\mathbf{Q} \mathbf{D} \mathbf{Q}^T) = \text{tr}(\mathbf{D}) = \sum_i d_i = \text{tr}(\bar{\mathbf{F}}^T \bar{\mathbf{F}}) \\ &= \|\bar{\mathbf{F}}\|_{\mathbf{F}}^2 = \sum_j \|\bar{\mathbf{F}}_{*j}\|^2,\end{aligned}\quad (\text{A.20})$$

where $\|\mathbf{A}\|_{\mathbf{F}}$ denotes the Frobenius norm of (Meyer 2000) \mathbf{A} . Hence, by placing observations where the Frobenius norm of the forward sensitivity matrix $\bar{\mathbf{F}}$ is a maximum, we can achieve the twin objective stated in the property A11.

Property A12: From remark 4.1, it is immediate that the matrix \mathbf{G} is also the Hessian of $J(\mathbf{c})$ at the minimum $\bar{\mathbf{c}}$.

Clearly, the condition number of \mathbf{G} controls the shape of the cost functional $J(\mathbf{c})$ around its minimum.

APPENDIX B

Factorization of $\mathbf{G} = \mathbf{F}^T \bar{\mathbf{H}} \mathbf{F}$

Recall from (3.14) that $\mathbf{G} = \mathbf{F}^T \bar{\mathbf{H}} \mathbf{F}$, where $\mathbf{F} = [\mathbf{U}, \mathbf{V}]$ and $\bar{\mathbf{H}} = \mathbf{D}_h^T(k) \mathbf{R}^{-1} \mathbf{D}_h(k)$. It is assumed that the noise covariance matrix \mathbf{R} (and hence \mathbf{R}^{-1}) is SPD. Accordingly, we can express \mathbf{R}^{-1} as the product of its Cholesky factors (Lewis et al. 2006, chapter 9) as

$$\mathbf{R}^{-1} = \mathbf{\Sigma}^T \mathbf{\Sigma}, \quad (\text{B.1})$$

where $\mathbf{\Sigma}$ is an upper triangular matrix.

Substituting, we obtain the Gramian structure

$$\mathbf{G} = \mathbf{F}^T \bar{\mathbf{H}} \mathbf{F} = (\mathbf{F}^T \mathbf{D}_h^T \mathbf{\Sigma}^T) (\mathbf{\Sigma} \mathbf{D}_h \mathbf{F}) = \bar{\mathbf{F}}^T \bar{\mathbf{F}}, \quad (\text{B.2})$$

where $\bar{\mathbf{F}} = \mathbf{\Sigma} \mathbf{D}_h(k) \mathbf{F}$.

In the first special case when $\mathbf{F} = \mathbf{U}$ (which occurs when $\delta \alpha = 0$), we get

$$\mathbf{G} = \mathbf{U}^T \bar{\mathbf{H}} \mathbf{U} = (\mathbf{U}^T \mathbf{D}_h^T \mathbf{\Sigma}^T) (\mathbf{\Sigma} \mathbf{D}_h \mathbf{U}) = \bar{\mathbf{U}}^T \bar{\mathbf{U}}, \quad (\text{B.3})$$

where $\bar{\mathbf{U}} = \mathbf{\Sigma} \mathbf{D}_h(k) \mathbf{U}$.

In the second special case when $\mathbf{F} = \mathbf{V}$ [which occurs when $\delta \mathbf{x}(0) = 0$], we get

$$\mathbf{G} = \mathbf{V}^T \bar{\mathbf{H}} \mathbf{V} = (\mathbf{V}^T \mathbf{D}_h^T \mathbf{\Sigma}^T) (\mathbf{\Sigma} \mathbf{D}_h \mathbf{V}) = \bar{\mathbf{V}}^T \bar{\mathbf{V}}, \quad (\text{B.4})$$

where $\bar{\mathbf{V}} = \mathbf{\Sigma} \mathbf{D}_h(k) \mathbf{V}$.

APPENDIX C

A Comparative Analysis

By way of comparing our approach with those in the adjoint sensitivity literature, we consider the adjoint sensitivity relation given in Eq. (8) in AH2007.

Solving Eq. (3.2) in this paper, it can be verified that the forward sensitivity matrix $\mathbf{U}(k)$ given by

$$\begin{aligned}\mathbf{U}(k) &= \mathbf{D}_M(k-1)\mathbf{D}_M(k-2)\dots\mathbf{D}_M(1)\mathbf{D}_M(0) \\ &= \mathbf{D}_M(k-1:0)\end{aligned}\quad (\text{C.1})$$

is indeed the same as resolvent $\mathbf{R}(k-1, 0)$ in the Eq. (2) in AH2007. Consequently, Eq. (2) in AH2007, in our notation, takes the form

$$\delta\mathbf{x}(k) = \mathbf{U}(k)\delta\mathbf{x}(0). \quad (\text{C.2})$$

The induced first variation δJ in J is then given by

$$\delta J = J[\mathbf{x}(k) + \delta\mathbf{x}(k)] - J[\mathbf{x}(k)] = \langle \nabla_{\mathbf{x}(k)} J, \delta\mathbf{x}(k) \rangle, \quad (\text{C.3})$$

where $\nabla_{\mathbf{x}(k)} J \in R^n$ is the gradient of $J[\mathbf{x}(k)]$ with respect to $\mathbf{x}(k)$

Combining (C.2) and (C.3) and using the adjoint property, from first principles it follows that the adjoint gradient is given by

$$\nabla_{\mathbf{x}(0)} J = \mathbf{U}^T(k)\nabla_{\mathbf{x}(k)} J, \quad (\text{C.4})$$

which is Eq. (8) in AH2007.

The analysis in AH2007 then proceeds to estimating the adjoint gradient using an ensemble approach. This is accomplished by building a linear regression between the ensemble of induced first variations $\{\delta J_i | 1 \leq i \leq M\}$ and the ensemble of the first variation in the initial conditions $\{\delta x_i(0) | 1 \leq i \leq M\}$, where M is the ensemble size. For further details, refer to AH2007.

However, in this paper we take the basic relation in (C.4) a step further by explicitly estimating $\nabla_{\mathbf{x}(k)} J$. From the structure of $J(\mathbf{c})$ in Eq. (2.6) of this paper, it can be verified that

$$\begin{aligned}-\nabla_{\mathbf{x}(k)} J &= \mathbf{D}_h^T(k)\mathbf{R}^{-1}\mathbf{e}(k) \\ &= \mathbf{D}_h^T(k)\mathbf{R}^{-1}\{\mathbf{z}(k) - \mathbf{h}[\mathbf{x}(k)]\} \\ &= \mathbf{D}_h^T(k)\mathbf{R}^{-1}\{\mathbf{h}[\bar{\mathbf{x}}(k)] - \mathbf{h}[\mathbf{x}(k)]\} \\ &= \mathbf{D}_h^T(k)\mathbf{R}^{-1}\mathbf{D}_h(k)\delta\mathbf{x}(k) \\ &= [\mathbf{D}_h^T(k)\mathbf{R}^{-1}\mathbf{D}_h(k)]\mathbf{U}(k)\delta\mathbf{x}(0).\end{aligned}\quad (\text{C.5})$$

Substituting Eq. (C.5) into (C.4), we get

$$-\nabla_{\mathbf{x}(0)} J = \mathbf{U}^T(k)[\mathbf{D}_h^T(k)\mathbf{R}^{-1}\mathbf{D}_h(k)]\mathbf{U}(k)\delta\mathbf{x}(0) = \mathbf{G}\delta\mathbf{x}(0), \quad (\text{C.6})$$

which is the equation in (4.1) for the special case when $\mathbf{x}(0)$ is varied and $\boldsymbol{\alpha}$ is held constant.

It is important to note that while all papers related to adjoint sensitivity analysis in the literature consider sensitivity only with respect to initial conditions, our work simultaneously deals sensitivity with respect to both initial conditions and parameters on the same footing.

REFERENCES

- Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, <https://doi.org/10.1175/2007MWR1904.1>.
- Bénard, M., 1900: Les tourbillons cellulaires dans une nappe liquide. *Rev. Gen. Sci. Pures Appl.*, **11**, 1261–1271, 1309–1328.
- , 1901: Les tourbillons cellulaires dans une nappe liquid transportant de la chaleur par convection en permanent. *Ann. Chim. Phys.*, **23**, 62–144.
- Berliner, L. M., Z. Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.*, **56**, 2536–2552, [https://doi.org/10.1175/1520-0469\(1999\)056<2536:SDFAWO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<2536:SDFAWO>2.0.CO;2).
- Bjerknes, J., and E. Palmén, 1937: Investigations of selected European cyclones by ascents. *Geofys. Publ.*, **12**, 1–62.
- Boussinesq, J., 1903: *Théorie Analytique de la Chaleur*. Gauthier-Villars, 670 pp.
- Burpee, R. W., J. L. Franklin, S. J. Lord, R. E. Tuleya, and S. D. Aberson, 1996: The impact of Omega dropwindsondes on operational hurricane track forecast models. *Bull. Amer. Meteor. Soc.*, **77**, 925–933, [https://doi.org/10.1175/1520-0477\(1996\)077<0925:TIOODO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0925:TIOODO>2.0.CO;2).
- Chandrasekhar, S., 1961: *Hydrodynamic and Hydromagnetic Stability*. Clarendon Press, 654 pp.
- Cochran, W. G., and G. M. Cox, 1992: *Experimental Designs*. 2nd ed. John Wiley and Sons, 640 pp.
- Eliassen, A., 1995: *Jacob Aall Bonnevie Bjerknes (1897–1975): Biographical Memoir*. National Academies Press, 21 pp.
- Evans, M. N., A. Kaplan, and M. A. Cane, 1998: Optimal sites for coral-based reconstruction of global sea surface temperature. *Paleoceanography*, **13**, 502–516, <https://doi.org/10.1029/98PA02132>.
- Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic–Dynamic Meteorology and Weather Analysis and Forecasting, Meteor. Monogr.*, No. 55, Amer. Meteor. Soc., 147–161, https://doi.org/10.1007/978-0-933876-68-2_7.
- Koch, S., M. Fengler, P. B. Chilson, K. L. Elmore, B. Argrow, D. L. Andra Jr., and T. Lindley, 2018: On the use of unmanned aircraft for sampling mesoscale phenomena in the preconvective boundary layer. *J. Atmos. Oceanic Technol.*, **35**, 2265–2288, <https://doi.org/10.1175/JTECH-D-18-0101.1>.
- Kotsuki, S. K., K. Kurosawa, and T. Miyoshi, 2019: On the properties of ensemble forecast sensitivity to observations. *Quart. J. Roy. Meteor. Soc.*, **145**, 1897–1914, <https://doi.org/10.1002/qj.3534>.
- Lakshmiarahan, S., and J. M. Lewis, 2010: Forward sensitivity based approach to dynamic data assimilation. *Adv. Meteor.*, **2010**, 375615, <https://doi.org/10.1155/2010/375615>.
- , —, and D. Phan, 2013: Data assimilation as a problem in optimal tracking: Application of Pontryagin’s minimum principle. *J. Atmos. Sci.*, **70**, 1257–1277, <https://doi.org/10.1175/JAS-D-12-0217.1>.

- , —, and R. Jabrzemski, 2017: *Forecast Error Correction Using Dynamic Data Assimilation*. Springer, 270 pp., <https://doi.org/10.1007/978-3-319-39997-3>.
- , —, and J. Hu, 2019a: Saltzman's model: Complete characterization of solution properties. *J. Atmos. Sci.*, **76**, 1587–1608, <https://doi.org/10.1175/JAS-D-17-0344.1>.
- , —, and —, 2019b: On controlling the shape of the cost functional in dynamic data assimilation: Guidelines for placement of observations—Part 1. University of Oklahoma School of Computer Science Tech. Rep., 39 pp.
- Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, <https://doi.org/10.1111/J.1600-0870.2004.00056.X>.
- , and Coauthors, 1999: The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Amer. Meteor. Soc.*, **80**, 1363–1384, [https://doi.org/10.1175/1520-0477\(1999\)080<1363:TNPEX>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<1363:TNPEX>2.0.CO;2).
- Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110, <https://doi.org/10.1111/j.1600-0870.1986.tb00459.x>.
- Lewis, J. M., and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309–322, <https://doi.org/10.3402/tellusa.v37i4.11675>.
- , S. Lakshmivarahan, and S. K. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, 654 pp.
- , —, and J. Hu, 2019: A criterion for choosing observation sites in data assimilation: Applied to Saltzman's convection model—Part 2. University of Oklahoma School of Computer Science Tech. Rep., 41 pp.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- , 1993: *The Essence of Chaos*. University of Washington, 227 pp.
- , and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, [https://doi.org/10.1175/1520-0469\(1998\)055<0399:OSFSWO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2).
- Majumdar, S. J., 2016: A review of targeted observations. *Bull. Amer. Meteor. Soc.*, **97**, 2287–2303, <https://doi.org/10.1175/BAMS-D-14-00259.1>.
- , and Coauthors, 2011: Targeted observations for improving numerical weather prediction: An overview. WMO Rep. WWRP/THORPEX 15, 37 pp., www.wmo.int/pages/prog/arep/wwrp/new/documents/THORPEX_No_15.pdf.
- Malkus, W. V. R., and G. Veronis, 1958: Finite amplitude cellular convection. *J. Fluid Mech.*, **4**, 225–260, <https://doi.org/10.1017/S0022112058000410>.
- Manohar, K., B. W. Brunton, J. N. Kutz, and S. L. Brunton, 2018: Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Syst. Mag.*, **38**, 63–86, <https://doi.org/10.1109/MCS.2018.2810460>.
- Meyer, C. D., 2000: *Matrix Analysis and Applied Linear Algebra*. SIAM, 718 pp.
- Narendra, K. S., and A. Annaswamy, 2005: *Stable Adaptive Systems*. Dover Publications, 496 pp.
- Rayleigh, L., 1916: Convection currents in a horizontal layer of fluid, when higher temperature is on the underside. *Philos. Mag.*, **32**, 529–546, <https://doi.org/10.1080/14786441608635602>.
- Saltzman, B., 1962: Finite amplitude free convection as an initial value problem—I. *J. Atmos. Sci.*, **19**, 329–341, [https://doi.org/10.1175/1520-0469\(1962\)019<0329:FAFCAA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1962)019<0329:FAFCAA>2.0.CO;2).
- Tolman, R. C., 2010: *Principles of Statistical Mechanics*. Dover Publications, 704 pp.
- Torn, R. D., and G. J. Hakim, 2008: Ensemble based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, <https://doi.org/10.1175/2007MWR2132.1>.