SENSITIVITY ANALYSIS IN NONLINEAR VARIATIONAL DATA ASSIMILATION: THEORETICAL ASPECTS AND APPLICATIONS

Dacian N. Daescu [1*] and Ionel M. Navon[†]

[*] *Portland State University, Portland, Oregon, U.S.A.*
[†] *Florida State University, Tallahassee, Florida, U.S.A.*

Abstract This chapter presents the mathematical framework to evaluate the sensitivity of a model forecast aspect to the input parameters of a nonlinear four-dimensional variational data assimilation system (4D-Var DAS): observations, prior state (background) estimate, and the error covariance specification. A fundamental relationship is established between the forecast sensitivity with respect to the information vector and the sensitivity with respect to the DAS representation of the information error covariance. Adjoint modeling is used to obtain first- and second-order derivative information and a reduced-order approach is formulated to alleviate the computational cost associated with the sensitivity estimation. Numerical results from idealized 4D-Var experiments performed with a global shallow water model are used to illustrate the theoretical concepts.

INTRODUCTION

Data assimilation techniques aim to produce an optimal estimate (analysis) of the state of the atmosphere by combining information extracted from a model representation of the atmospheric dynamics, observational data, and error statistics [1, 2, 3, 4]. In the recent years, the amount of data provided by global observatory network has increased at a fast pace as a result of advanced measurement and data processing capabilities such as the ability to infer atmospheric parameters from radiance measurements provided by the satellite sounding instruments [5]. Suboptimal weighting of the information provided by models and measurements poses a fundamental limitation on the performance of data assimilation systems (DAS) and it is unanimously accepted that a major source of uncertainty in the analyses and forecasts is due to the practical difficulty of providing accurate estimates to the error statistics [6]. A significant amount of research is focused on modeling the observation and background error covariance matrices [7, 8, 9]. Data thinning, super-obbing, and error variance inflation are the traditional procedures implemented at numerical weather prediction (NWP) centers to address the information-redundancy and to alleviate the impact of observational error correlations that are not represented in the DAS [10].

The development of efficient methodologies to quantify the contribution of each observing system component to the analysis and forecast error reduction is a practical necessity and focus of research at atmospheric research centers worldwide [11, 12, 13, 14, 15, 16]. Novel sensitivity techniques are needed to provide guidance to error covariance tuning procedures and *a priori* estimates of the analysis and forecast impact due to the variations in the specification of the input error statistics. These techniques should be feasible for practical implementation and effective in identifying the DAS input components whose improved estimates of the error statistics will render increased benefits to the forecasts.

A proper understanding of how uncertainties in the specification of the input error statistics will impact the analysis and forecasts may be achieved by performing a sensitivity analysis with respect to an augmented set of parameters in the DAS that include observations, background estimate, and the associated error covariances. The present work provides a review of the mathematical aspects and recent developments in the formulation of sensitivity analysis methods in four-dimensional variational (4D-Var) data assimilation. Results from idealized data assimilation experiments are used to illustrate some of the theoretical concepts and further research directions for practical applications are discussed.

[1]**Correspondence Author:** Dacian N. Daescu, Portland State University, P.O. Box 751 Portland, OR 97207, USA; E-mail: daescu@pdx.edu

4D-VAR DATA ASSIMILATION

The weak-constraint four-dimensional variational data assimilation (w4D-Var) provides an estimate (analysis) $\mathbf{x}^a$ to the atmospheric state $\mathbf{x}^t$ by solving a generalized nonlinear least-squares optimization problem [1, 2, 17]

$$\begin{aligned} J &= J^b + J^o + J^q \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ &+ \frac{1}{2}\sum_{k=0}^{N}[\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k]^{\mathrm{T}}\mathbf{R}_k^{-1}[\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k] \\ &+ \frac{1}{2}\sum_{k=1}^{N}[\mathbf{x}_k - M_{k-1,k}(\mathbf{x}_{k-1})]^{\mathrm{T}}\mathbf{Q}_k^{-1}[\mathbf{x}_k - M_{k-1,k}(\mathbf{x}_{k-1})] \end{aligned} \tag{1}$$

with respect to the time-distributed (four dimensional) state vector $\mathbf{x}^{\mathrm{T}} = [\mathbf{x}_0^{\mathrm{T}}\,\mathbf{x}_1^{\mathrm{T}}\,\ldots\,\mathbf{x}_N^{\mathrm{T}}]$ in the analysis interval $[t_0,\,t_N]$. The cost (1) incorporates information from a prior (background) estimate $\mathbf{x}_0^b \in \mathbb{R}^n$ to the true initial state $\mathbf{x}_0^t$, time-distributed observational data $\mathbf{y}_k \in \mathbb{R}^{p_k}, k = 0 : N$, and the atmospheric model M. The analysis $\mathbf{x}^a$ is closely determined by the specification of the weight matrices $\mathbf{B}, \mathbf{R}_k$, and $\mathbf{Q}_k$ that are representations in the DAS of the covariance matrices $\mathbf{B}^t, \mathbf{R}_k^t$, and $\mathbf{Q}_k^t$ of the background errors $\boldsymbol{\epsilon}_0^b = \mathbf{x}_0^b - \mathbf{x}_0^t$, observational errors $\boldsymbol{\epsilon}_k^o = \mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k^t)$, and model errors $\boldsymbol{\epsilon}_k^q = \mathbf{x}_k^t - M_{k-1,k}(\mathbf{x}_{k-1}^t)$, respectively. In the formulation (1) it is assumed that the information errors $\boldsymbol{\epsilon}_0^b, \boldsymbol{\epsilon}_k^o, \boldsymbol{\epsilon}_k^q$ are uncorrelated and additional simplifying assumptions are necessary to achieve a practical implementation (Rabier 2005). The strong-constraint 4D-Var ignores the model errors and, by imposing the model equations

$$\mathbf{x}_k = M_{k-1,k}(\mathbf{x}_{k-1}), \quad k = 1 : N \tag{2}$$

as a strong constraint, the analysis state at the initial time t_0 is obtained by solving an initial-condition optimization problem with the cost functional defined as

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\left[\mathbf{h}(\mathbf{x}_0) - \mathbf{y}\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{h}(\mathbf{x}_0) - \mathbf{y}\right] \tag{3}$$

$$\mathbf{x}_0^a = \mathrm{Arg}\ \min J \tag{4}$$

The operator $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$ incorporates the nonlinear model to advance the initial state to the observation time,

$$\mathbf{h}(\mathbf{x}_0) - \mathbf{y} = \begin{bmatrix} \mathbf{h}_0(\mathbf{x}_0) - \mathbf{y}_0 \\ \mathbf{h}_1(\mathbf{x}_1) - \mathbf{y}_1 \\ \vdots \\ \mathbf{h}_N(\mathbf{x}_N) - \mathbf{y}_N \end{bmatrix} \in \mathbb{R}^p \tag{5}$$

where $p = p_0 + p_1 + \ldots p_N$ denotes the dimension of the time-distributed observation vector $\mathbf{y}$, and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & 0 & \ldots & 0 \\ 0 & \mathbf{R}_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{R}_N \end{bmatrix} \in \mathbb{R}^{p\times p} \tag{6}$$

is the block diagonal observation error covariance matrix. In practice, an approximate solution to the nonlinear optimization problem (3-4) is obtained using an iterative minimization algorithm.

First-Order Derivative Information

The gradient of the cost functional is expressed as

$$\nabla_{\mathbf{x}_0} J = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{h}(\mathbf{x}_0) - \mathbf{y}\right] \tag{7}$$

where the operator $\mathbf{H}$ incorporates both the linearized observation operator

$$\mathbf{H}_k = \left[\frac{\partial \mathbf{h}_k}{\partial \mathbf{x}}\right]_{|\mathbf{x}=\mathbf{x}_k} \in \mathbb{R}^{p_k \times n},\ \ k = 0 : N \tag{8}$$

and the tangent linear model (TLM) associated with the nonlinear forecast model equations (2)

$$\mathbf{M}_{0,k} = \mathbf{M}_{k-1,k} \ldots \mathbf{M}_{1,2}\mathbf{M}_{0,1}(\mathbf{x}_0) \tag{9}$$

and is defined as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{0,N} \end{bmatrix} \in \mathbb{R}^{p \times n} \tag{10}$$

A computationally feasible approach to evaluate the gradient (7) may be implemented by developing the adjoint $\mathbf{M}^{\mathrm{T}}$ of the tangent linear model. Backward integration of the adjoint equations

$$\boldsymbol{\lambda}_N = \mathbf{H}_N^{\mathrm{T}}\mathbf{R}_N^{-1}[\mathbf{h}_N(\mathbf{x}_N) - \mathbf{y}_N] \tag{11}$$

$$\boldsymbol{\lambda}_k = \mathbf{M}_{k,k+1}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1} + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}[\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k], \quad k = N-1 : -1 : 0 \tag{12}$$

provides the gradient (7) as

$$\nabla_{\mathbf{x}_0} J = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \boldsymbol{\lambda}_0 \tag{13}$$

Large-scale optimization algorithms that rely only on first-order derivative information such as the quasi-Newton limited-memory BFGS method [19] may be used to provide an iterative solution to (3)-(4),

$$\mathbf{x}_0^{a,(i+1)} = \mathbf{x}_0^{a,(i)} + \alpha_i \mathbf{d}^{(i)}, \quad i = 0, 1, \ldots \tag{14}$$

where $\mathbf{x}_0^{a,(0)} = \mathbf{x}_0^b$ is the initial estimate, $\mathbf{d}^{(i)}$ is a descent direction, and $\alpha_i > 0$ is a properly selected step length [20, 21].

The Incremental Approach

The incremental 4D-Var introduced by Courtier et al. [22] relies on successive quadratic approximations (outer-loop iteration) to the nonlinear problem (3) and provides a feasible approach for operational implementation of the 4D-Var scheme [18]. The outer-loop iteration is expressed as

$$\mathbf{x}_0^{a,(i+1)} = \mathbf{x}_0^{a,(i)} + \delta\mathbf{x}_0^{a,(i)},\ i = 0, 1, \ldots \tag{15}$$

where $\mathbf{x}_0^{a,(0)} = \mathbf{x}_0^b$ is the prior estimate to the initial condition. The analysis increment $\delta\mathbf{x}_0^{a,(i)}$ is obtained by solving the quadratic optimization problem

$$\begin{aligned} \widetilde{J}(\delta\mathbf{x}_0^{(i)}) &= \frac{1}{2}\left[\delta\mathbf{x}_0^{(i)} - \left(\mathbf{x}_0^b - \mathbf{x}_0^{a,(i)}\right)\right]^{\mathrm{T}}\mathbf{B}^{-1}\left[\delta\mathbf{x}_0^{(i)} - \left(\mathbf{x}_0^b - \mathbf{x}_0^{a,(i)}\right)\right] \\ &+ \frac{1}{2}\left[\mathbf{H}^{a,(i)}\delta\mathbf{x}_0^{(i)} - \left(\mathbf{y} - \mathbf{h}(\mathbf{x}^{a,(i)})\right)\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{H}^{a,(i)}\delta\mathbf{x}_0^{(i)} - \left(\mathbf{y} - \mathbf{h}(\mathbf{x}^{a,(i)})\right)\right] \\ \delta\mathbf{x}_0^{a,(i)} &= \mathrm{Arg\ min}\,\widetilde{J} \end{aligned} \tag{16}$$

where $\mathbf{H}^{a,(i)}$ denotes the linearized observational operator (10) evaluated at $\mathbf{x}_0^{a,(i)}$, $\mathbf{H}^{a,(i)} = \mathbf{H}(\mathbf{x}_0^{a,(i)})$. The optimal analysis increment (16) is expressed as

$$\delta \mathbf{x}_0^{a,(i)} = \mathbf{x}_0^b - \mathbf{x}_0^{a,(i)} + \mathbf{K}^{a,(i)} \left[\mathbf{y} - \mathbf{h}(\mathbf{x}^{a,(i)}) - \mathbf{H}^{a,(i)} \left(\mathbf{x}_0^b - \mathbf{x}_0^{a,(i)} \right) \right] \tag{17}$$

where $\mathbf{K}^{a,(i)}$ denotes the optimal gain matrix (DAS operator) evaluated at $\mathbf{x}_0^{a,(i)}$,

$$\mathbf{K}^{a,(i)} = \mathbf{B}\mathbf{H}^{a,(i)^{\mathrm{T}}} \left[\mathbf{H}^{a,(i)}\mathbf{B}\mathbf{H}^{a,(i)^{\mathrm{T}}} + \mathbf{R} \right]^{-1} = \left[\mathbf{B}^{-1} + \mathbf{H}^{a,(i)^{\mathrm{T}}}\mathbf{R}^{-1}\mathbf{H}^{a,(i)} \right]^{-1} \mathbf{H}^{a,(i)^{\mathrm{T}}}\mathbf{R}^{-1} \tag{18}$$

By replacing (17) in (15), the outer-loop iteration is expressed as

$$\mathbf{x}_0^{a,(i+1)} = \mathbf{x}_0^b + \mathbf{K}^{a,(i)} \left[\mathbf{y} - \mathbf{h}(\mathbf{x}^{a,(i)}) - \mathbf{H}^{a,(i)} \left(\mathbf{x}_0^b - \mathbf{x}_0^{a,(i)} \right) \right] \tag{19}$$

Evaluation of the right side term in (19) requires the solution to a linear system and is performed using an iterative algorithm (inner-loop iteration) such as the conjugate gradient method. The implementation of the incremental 4D-Var algorithm and issues related with the convergence of the outer loop iteration are discussed in [23], [24], [25]. Often, for practical reasons, a single outer loop iteration is implemented and the analysis state is obtained as

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \mathbf{K} \left[\mathbf{y} - \mathbf{h}(\mathbf{x}_0^b) \right] \tag{20}$$

where the DAS operator $\mathbf{K}$ is defined as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{b^{\mathrm{T}}} \left[\mathbf{H}^b\mathbf{B}\mathbf{H}^{b^{\mathrm{T}}} + \mathbf{R} \right]^{-1} = \left[\mathbf{B}^{-1} + \mathbf{H}^{b^{\mathrm{T}}}\mathbf{R}^{-1}\mathbf{H}^b \right]^{-1} \mathbf{H}^{b^{\mathrm{T}}}\mathbf{R}^{-1} \tag{21}$$

SENSITIVITY ANALYSIS IN NONLINEAR 4D-VAR

Fundamental aspects of the mathematical theory and practical aspects of sensitivity analysis for nonlinear dynamical systems are discussed in the books of Cacuci [26], Cacuci et al. [27], Marchuk et al. [28], and Saltelli et al. [29]. A theoretical framework to sensitivity analysis in VDA is presented by Le Dimet et al. [30]. For atmospheric data assimilation applications, the forecast score (quantity of interest) is typically defined as a short-range forecast error measure

$$e(\mathbf{x}_0) = (\mathbf{x}_f - \mathbf{x}_f^v)^{\mathrm{T}}\mathbf{E}(\mathbf{x}_f - \mathbf{x}_f^v) \tag{22}$$

where $\mathbf{x}_f \in \mathbb{R}^n$ is the model forecast at verification time t_f initiated at t_0 from $\mathbf{x}_0$, $\mathbf{x}_f = M_{0,t_f}(\mathbf{x}_0)$, $\mathbf{x}_f^v$ is the verifying analysis at t_f and serves as a proxy to the true state $\mathbf{x}_f^t$, and $\mathbf{E}$ is a diagonal matrix of weights that gives (22) units of energy per unit mass.

The first-order variation in the functional $e(\mathbf{x}_0^a)$ induced by an analysis variation $\delta\mathbf{x}_0^a$ is expressed as

$$\delta e = \langle \nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a), \delta\mathbf{x}_0^a \rangle_{\mathbb{R}^n} \tag{23}$$

where $\nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a)$ denotes the forecast sensitivity (gradient) to the analysis and is obtained using a backward adjoint model integration from t_f to t_0 along the analysis trajectory

$$\nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) = \left[\mathbf{M}_{0,t_f}^a \right]^{\mathrm{T}} \mathbf{E}(\mathbf{x}_f^a - \mathbf{x}_f^v) \tag{24}$$

The forecast sensitivity to the DAS input parameters $\mathbf{y}, \mathbf{R}, \mathbf{x}_0^b, \mathbf{B}$ is formulated in the context of the 4D-Var optimization problem (3-4), $\mathbf{x}_0^a = \mathbf{x}_0^a(\mathbf{y}, \mathbf{R}, \mathbf{x}_0^b, \mathbf{B})$. A first study was presented by Daescu [31] where the 4D-Var sensitivity equations were derived from the first order optimality condition to (3) using the implicit function theorem and matrix differentiation calculus. In this section a derivation of the sensitivity equations in nonlinear 4D-Var is provided using first-order perturbation theory. Computational aspects are addressed including a reduced-order method to sensitivity estimation. The practical ability to perform error covariance sensitivity analysis relies on the close relationships that may be established between the forecast sensitivity to observations, background, and the associated error covariance models $\mathbf{R}$ and $\mathbf{B}$, respectively.

4D-Var Sensitivity Equations

For notational convenience and to provide a compact derivation of the 4D-Var sensitivity equations, let

$$\mathbf{w} = \begin{bmatrix} \mathbf{x}_0^b \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{n+p} \tag{25}$$

denote the information vector,

$$\mathbf{\Gamma} : \mathbb{R}^n \to \mathbb{R}^{n+p}, \quad \mathbf{\Gamma}(\mathbf{x}_0) = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{h}(\mathbf{x}_0) \end{bmatrix} \tag{26}$$

denote the nonlinear operator that maps the initial state into the information space, and

$$\mathbf{W} = \begin{bmatrix} \mathbf{B} & \mathbf{0}_{n\times p} \\ \mathbf{0}_{p\times n} & \mathbf{R} \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)} \tag{27}$$

denote the DAS model of the information error covariance. With this notation, the 4D-Var cost functional (3) is expressed as

$$J(\mathbf{x}_0) = \frac{1}{2}\left[\mathbf{\Gamma}(\mathbf{x}_0) - \mathbf{w}\right]^{\mathrm{T}} \mathbf{W}^{-1} \left[\mathbf{\Gamma}(\mathbf{x}_0) - \mathbf{w}\right] \tag{28}$$

An implicit relationship $\mathbf{x}_0^a = \mathbf{x}_0^a(\mathbf{w}, \mathbf{W})$ between the analysis $\mathbf{x}_0^a$ and the DAS input parameters $\mathbf{w}$ and $\mathbf{W}$ is established from the first-order optimality condition

$$\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^a) = \mathbf{0} \Leftrightarrow \mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right] = \mathbf{0} \tag{29}$$

where

$$\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a) = \left[\frac{\partial \mathbf{\Gamma}}{\partial \mathbf{x}_0}\right]_{|\mathbf{x}_0 = \mathbf{x}_0^a} = \begin{bmatrix} \mathbf{I}_{n\times n} \\ \mathbf{H}^a \end{bmatrix} \in \mathbb{R}^{(n+p)\times n} \tag{30}$$

denotes the Jacobian matrix of $\mathbf{\Gamma}$ evaluated at $\mathbf{x}_0^a$. The forecast sensitivity to $\mathbf{w}$ and $\mathbf{W}$ is obtained by relating the first-order variation $\delta\mathbf{x}_0^a$ with the parameter variations $\delta\mathbf{w}$ and $\delta\mathbf{W}$, respectively. A relationship is established from (29) as

$$\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a)\right]\delta\mathbf{x}_0^a - \mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}\delta\mathbf{w} - \mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}[\delta\mathbf{W}]\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right] = \mathbf{0} \tag{31}$$

where

$$\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a) = \left(\frac{\partial[\nabla_{\mathbf{x}_0} J]}{\partial \mathbf{x}_0}\right)_{|\mathbf{x}_0 = \mathbf{x}_0^a} \in \mathbb{R}^{n\times n} \tag{32}$$

denotes the Hessian matrix of the 4D-Var cost functional evaluated at $\mathbf{x}_0^a(\mathbf{w}, \mathbf{W})$. The identity

$$\delta[\mathbf{W}^{-1}] = -\mathbf{W}^{-1}[\delta\mathbf{W}]\mathbf{W}^{-1} \tag{33}$$

was used to derive the last term in the left side of (31). By taking the inner product of (31) with an arbitrary vector $\boldsymbol{\eta} \in \mathbb{R}^n$ we obtain

$$\left\langle \boldsymbol{\eta}, \left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a)\right]\delta\mathbf{x}_0^a \right\rangle_{\mathbb{R}^n} = \left\langle \boldsymbol{\eta}, \mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}\delta\mathbf{w} \right\rangle_{\mathbb{R}^n} + \left\langle \boldsymbol{\eta}, \mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}[\delta\mathbf{W}]\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right] \right\rangle_{\mathbb{R}^n} \tag{34}$$

Taking into accont that the Hessian $\nabla^2 J$ and the covariance model $\mathbf{W}$ are symmetric matrices and with the aid of a few linear algebra operations, (34) may be expressed as

$$\left\langle \left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a)\right]\boldsymbol{\eta}, \delta\mathbf{x}_0^a \right\rangle_{\mathbb{R}^n} = \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}, \delta\mathbf{w} \right\rangle_{\mathbb{R}^{n+p}} + \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}, [\delta\mathbf{W}]\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right] \right\rangle_{\mathbb{R}^{n+p}} \tag{35}$$

Therefore, by defining $\boldsymbol{\eta}$ as the solution to the linear system

$$\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a)\right]\boldsymbol{\eta} = \nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) \tag{36}$$

from (23), (35), and (36) the first-order variation in the functional aspect is expressed as

$$\delta e = \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}, \delta\mathbf{w}\right\rangle_{\mathbb{R}^{n+p}} + \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}, [\delta\mathbf{W}]\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right]\right\rangle_{\mathbb{R}^{n+p}} \quad (37)$$

We recall that the Frobenius inner product of two matrices of the same order $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\langle \mathbf{X}, \mathbf{Y}\rangle = Tr\left[\mathbf{X}\mathbf{Y}^{\mathrm{T}}\right] \quad (38)$$

where Tr denotes the matrix trace operator. The right side of (37) may be expressed as

$$\delta e = \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}, \delta\mathbf{w}\right\rangle_{\mathbb{R}^{n+p}} + \left\langle \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right]^{\mathrm{T}}\mathbf{W}^{-1}, \delta\mathbf{W}\right\rangle_{\mathbb{R}^{(n+p)\times(n+p)}} \quad (39)$$

and equation (39) provides the explicit relationship between the first-order forecast variation δe and the variations $\delta\mathbf{w}$ and $\delta\mathbf{W}$ in the DAS input parameters. The forecast sensitivity to the information vector $\mathbf{w}$ is expressed as

$$\nabla_{\mathbf{w}} e(\mathbf{x}_0^a) = \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}_0}(\mathbf{x}_0^a)\boldsymbol{\eta} \in \mathbb{R}^{n+p} \quad (40)$$

From the first-order optimality condition (29), it is noticed that for any forecast aspect the sensitivity to information $\nabla_{\mathbf{w}} e(\mathbf{x}_0^a)$ and the analysis fit to information $\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}$ are orthogonal vectors in $\mathbb{R}^{n+p}$,

$$\langle \nabla_{\mathbf{w}} e(\mathbf{x}_0^a), \mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\rangle_{\mathbb{R}^{n+p}} = \boldsymbol{\eta}^{\mathrm{T}}\mathbf{\Gamma}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{x}_0^a)\mathbf{W}^{-1}\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right] = 0 \quad (41)$$

From (39), a fundamental relationship is established between the forecast sensitivity to the information vector and the forecast sensitivity to the covariance model $\mathbf{W}$ that is expressed as the rank-one matrix

$$\frac{\partial e(\mathbf{x}_0^a)}{\partial \mathbf{W}} = \nabla_{\mathbf{w}} e(\mathbf{x}_0^a)\left[\mathbf{\Gamma}(\mathbf{x}_0^a) - \mathbf{w}\right]^{\mathrm{T}}\mathbf{W}^{-1} \in \mathbb{R}^{(n+p)\times(n+p)} \quad (42)$$

Taking into account the notational convenience (25-27) and (30), the equations of the forecast sensitivity to the DAS input $(\mathbf{y}, \mathbf{R})$ and $(\mathbf{x}_0^b, \mathbf{B})$ are as follows:

$$\begin{aligned}
\nabla_{\mathbf{y}} e(\mathbf{x}_0^a) &= \mathbf{R}^{-1}\mathbf{H}^a\boldsymbol{\eta} \in \mathbb{R}^p && (43)\\
\frac{\partial e(\mathbf{x}_0^a)}{\partial \mathbf{R}} &= \nabla_{\mathbf{y}} e(\mathbf{x}_0^a)[\mathbf{R}^{-1}(\mathbf{h}(\mathbf{x}_0^a) - \mathbf{y})]^{\mathrm{T}} \in \mathbb{R}^{p\times p} && (44)\\
\nabla_{\mathbf{x}_0^b} e(\mathbf{x}_0^a) &= \mathbf{B}^{-1}\boldsymbol{\eta} \in \mathbb{R}^n && (45)\\
\frac{\partial e(\mathbf{x}_0^a)}{\partial \mathbf{B}} &= \nabla_{\mathbf{x}_0^b} e(\mathbf{x}_0^a)[\mathbf{B}^{-1}(\mathbf{x}_0^a - \mathbf{x}_0^b)]^{\mathrm{T}} \in \mathbb{R}^{n\times n} && (46)
\end{aligned}$$

From (44) and (46) it is noticed that evaluation and storage of only a few vectors is necessary to capture the information content of the forecast **R**- and **B**-sensitivities. The first order optimality condition (29) provides the relationship

$$\mathbf{B}^{-1}(\mathbf{x}_0^a - \mathbf{x}_0^b) = \mathbf{H}^{a\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - \mathbf{h}(\mathbf{x}_0^a)] \quad (47)$$

such that a mathematically equivalent formulation to the **B**-sensitivity equation (46) may be obtained as

$$\frac{\partial e(\mathbf{x}_0^a)}{\partial \mathbf{B}} = \nabla_{\mathbf{x}_0^b} e(\mathbf{x}_0^a)\left\{\mathbf{H}^{a\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - \mathbf{h}(\mathbf{x}_0^a)]\right\}^{\mathrm{T}} \quad (48)$$

Second-Order Derivative Information

The sensitivity analysis of a nonlinear 4D-Var DAS requires the solution to the linear system (36) that involves the Hessian of the cost functional (3). Second-order derivative information may be obtained by developing a second-order adjoint (SOA) model [32]. The product between the Hessian matrix of the cost functional and a user-defined vector $\mathbf{v} \in \mathbb{R}^n$ may be evaluated using a forward over reverse procedure that requires the linearization of the nonlinear model equations (2) and of the first-order adjoint equations (11)-(13) with respect to the state and adjoint variables. Assuming, for simplicity, that the observation operator $\mathbf{h}_k$ is linear, $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}_k\mathbf{x}_k$, the SOA model equations are espressed as

$$\boldsymbol{\mu}_0 = \mathbf{v} \tag{49}$$

$$\boldsymbol{\mu}_k = \mathbf{M}_{k-1,k}\boldsymbol{\mu}_{k-1}, k = 1 : N \tag{50}$$

$$\boldsymbol{\xi}_N = \mathbf{H}_N^{\mathrm{T}}\mathbf{R}_N^{-1}\mathbf{H}_N\boldsymbol{\mu}_N \tag{51}$$

$$\boldsymbol{\xi}_k = \mathbf{M}_{k,k+1}^{\mathrm{T}}\boldsymbol{\xi}_{k+1} + \partial_{\mathbf{x}_k}\left[\mathbf{M}_{k,k+1}^{\mathrm{T}}\bar{\boldsymbol{\lambda}}_{k+1}\right]\boldsymbol{\mu}_k + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k\boldsymbol{\mu}_k, \quad k = N-1 : -1 : 0 \tag{52}$$

In the equation above, $\partial_{\mathbf{x}_k}\left[\mathbf{M}_{k,k+1}^{\mathrm{T}}\bar{\boldsymbol{\lambda}}_{k+1}\right]$ is a symmetric matrix that incorporates second-order $\mathbf{x}_k$-derivatives of the model $M_{k,k+1}(\mathbf{x}_k)$ and the notation $\bar{\boldsymbol{\lambda}}_{k+1}$ indicates that the first order adjoint variables $\boldsymbol{\lambda}_{k+1}$ are treated as constants during this differentiation. The integration of the SOA model along the analysis trajectory provides the Hessian/vector product as

$$\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0}J(\mathbf{x}_0^a)\right]\mathbf{v} = \mathbf{B}^{-1}\mathbf{v} + \boldsymbol{\xi}_0 \tag{53}$$

The development of a SOA model is demanding task and requires significant efforts in the software development as well as an increased number of floating point operations and memory storage. In practice, automatic differentiation tools may be used to facilitate the implementation of the SOA model [33], [34]. The second-order derivative information allows the implementation of Hessian-free truncated Newton (HFTN) methods for large-scale optimization [35], [36] and of hybrid methods that aim to improve the convergence of the iterative solution to (4) by dynamically interlacing inexpensive L-BFGS iterations with fast convergent HFTN iterations [37], [38]. Having available the Hessian/vector products, the evaluation of the vector $\boldsymbol{\eta}$ may be performed using an iterative solver such as the conjugate gradient (CG) method. Each iteration of the CG solver requires an additional forward integration of the TLM equations (49-50) with storage of the state trajectory $\mathbf{x}_k, k = 1 : N$ and tangent variables $\boldsymbol{\mu}_k, k = 1 : N$, and followed by the backward integration of the first order adjoint equations (11-12) and of the SOA equations (51-52) to obtain the Hessian/vector product (53). The relative ratio r between the CPU required by a SOA model integration and the CPU of a forward model integration (2) is typically in the range $7 \leq r \leq 10$ and may increase due to additional costs entailed by the implementation of checkpointing schemes and data manipulation [34].

To date, the increased computational burden has prevented the practical implementation of sensitivity analysis in nonlinear 4D-Var. Sensitivity studies in atmospheric data assimilation have been mainly considered for a linear analysis scheme (20) and have been used to assess the forecast sensitivity to observations $\mathbf{y}$ and, to a much smaller extent, to the specification of the background estimate $\mathbf{x}_0^b$. In the linear context, the evaluation of the sensitivity of a forecast aspect $e(\mathbf{x}_0^a)$ with respect to observations and to the background estimate was considered by Baker and Daley [39] for applications to observation targeting in NWP. For the analysis equation (20), the forecast sensitivity to observations is expressed as [39]

$$\nabla_{\mathbf{y}}e(\mathbf{x}_0^a) = \mathbf{K}^{\mathrm{T}}\nabla_{\mathbf{x}_0}e(\mathbf{x}_0^a) \tag{54}$$

where

$$\mathbf{K}^{\mathrm{T}} = \left[\mathbf{H}^b\mathbf{B}{\mathbf{H}^b}^{\mathrm{T}} + \mathbf{R}\right]^{-1}\mathbf{H}^b\mathbf{B} = \mathbf{R}^{-1}\mathbf{H}^b\left[\mathbf{B}^{-1} + {\mathbf{H}^b}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}^b\right]^{-1} \tag{55}$$

is the adjoint (transpose) of the DAS operator $\mathbf{K}$ (21). It is noticed that the computational cost to evaluate the observation sensitivity (54) is roughly equivalent with the computational cost to obtain the analysis (20). First applications of the error-covariance sensitivity analysis in operational 3D-Var and 4D-Var DAS with a single outer-loop iteration were presented by Daescu and Todling [40] and Daescu and Langland [41]. A study on observation sensitivity and observation impact calculations in a variational DAS implementing multiple outer loop iterations (19) is provided by Trémolet [42]. A computationally feasible approach to sensitivity analysis in nonlinear 4D-Var may be obtained by formulating a reduced-order optimization problem by projection on a low-dimensional subspace. A reduced-order approach to sensitivity estimation and the selection of a projection operator that incorporates in a consistent fashion information pertinent to the 4D-Var iteration and forecast sensitivity to the initial condition is presented in the next section.

Reduced-Order 4D-Var Sensitivity Estimation

The 4D-Var sensitivity equations (43-46) are derived from the optimality condition (29) whereas in the practical implementation the minimization to the cost functional (3) is terminated when the gradient satisfies a specified convergence criteria or simply after a prescribed number of iterations. An estimate of the observation sensitivity that is consistent with the data assimilation process may be obtained by implementing the adjoint of the minimization algorithm [43]. In this section a reduced-order approach to sensitivity estimation is formulated using information gathered in the 4D-Var minimization of the cost functional (3) to identify an appropriate low-order subspace.

In a general framework, the reduced-order 4D-Var is formulated by projecting the initial-condition increment $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^b$ into a low-order control space of dimension $m \ll n$ [44, 45],

$$\mathbf{\Pi}\delta\mathbf{x}_0 = \mathbf{\Psi}\boldsymbol{\zeta} = \sum_{j=1}^{m} \zeta_j \boldsymbol{\psi}_j \tag{56}$$

where the matrix $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_m] \in \mathbb{R}^{n\times m}$ incorporates the orthonormal vectors of the reduced-space basis $\boldsymbol{\psi}_j$, $j = 1:m$, $\mathbf{\Pi} = \mathbf{\Psi}\mathbf{\Psi}^{\mathrm{T}}$ denotes the projection operator from the full space onto the reduced-space, and $\boldsymbol{\zeta} = (\zeta_1, \ldots \zeta_m)^{\mathrm{T}} \in \mathbb{R}^m$ denotes the coordinates vector in the reduced space

$$\boldsymbol{\zeta} = \mathbf{\Psi}^{\mathrm{T}}\delta\mathbf{x}_0 \tag{57}$$

The reduced-order 4D-Var approach searches for the optimal coefficients $\boldsymbol{\zeta}^a \in \mathbb{R}^m$ by solving a low-dimensional optimization problem with the cost functional defined as

$$\widehat{J}(\boldsymbol{\zeta}) = J(\mathbf{x}_0^b + \mathbf{\Psi}\boldsymbol{\zeta}) \tag{58}$$

$$\boldsymbol{\zeta}^a = \mathrm{Arg\ min}\,\widehat{J} \tag{59}$$

and an approximation to the analysis (4) is obtained as

$$\mathbf{x}_0^a \approx \mathbf{x}_0^b + \mathbf{\Psi}\boldsymbol{\zeta}^a \tag{60}$$

The gradient and the Hessian matrix of the cost functional (58) are expressed as [44]

$$\nabla_{\boldsymbol{\zeta}}\widehat{J} = \mathbf{\Psi}^{\mathrm{T}}\nabla_{\mathbf{x}_0}J \tag{61}$$

$$\nabla^2_{\boldsymbol{\zeta}\boldsymbol{\zeta}}\widehat{J} = \mathbf{\Psi}^{\mathrm{T}}\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0}J\right]\mathbf{\Psi} \tag{62}$$

and in this fashion the large-scale (n-dimensional) linear system (36) is replaced by the low-order (m-dimensional) linear system

$$\left[\nabla^2_{\boldsymbol{\zeta}\boldsymbol{\zeta}}\widehat{J}(\boldsymbol{\zeta}^a)\right]\widehat{\boldsymbol{\eta}} = \mathbf{\Psi}^{\mathrm{T}}\nabla_{\mathbf{x}_0}e(\mathbf{x}_0^a) \tag{63}$$

The reduced-order Hessian matrix $\nabla^2_{\boldsymbol{\zeta}\boldsymbol{\zeta}}\widehat{J}$ is positive definite, provided that the full Hessian matrix $\nabla^2_{\mathbf{x}_0\mathbf{x}_0}J$ is positive definite, and the reduced-order approach provides an estimate to the vector $\boldsymbol{\eta}$ (36) as

$$\boldsymbol{\eta} \approx \mathbf{\Psi}\widehat{\boldsymbol{\eta}} \tag{64}$$

It is noticed that if the approximation (60) is exact then the reduced- order 4D-Var sensitivity estimation relies on the reduced-rank approximation to the inverse Hessian matrix

$$\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0}J(\mathbf{x}_0^a)\right]^{-1} \approx \mathbf{\Psi}\left[\mathbf{\Psi}^{\mathrm{T}}\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0}J(\mathbf{x}_0^a)\right]\mathbf{\Psi}\right]^{-1}\mathbf{\Psi}^{\mathrm{T}} \tag{65}$$

and $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}} \in \mathbb{R}^n$ is the solution to

$$\boldsymbol{\Pi}\left[\nabla^2_{\mathbf{x}_0\mathbf{x}_0} J(\mathbf{x}_0^a)\right] \boldsymbol{\Psi}\widehat{\boldsymbol{\eta}} = \boldsymbol{\Pi}\nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) \tag{66}$$

Low-rank approximations based on the leading eigenvectors of the Hessian matrix and the BFGS method are discussed in references [46, 47]. A computationally efficient approach for selecting a reduced-space that is consistent with the 4D-Var minimization process is presented below.

Reduced-Order Subspace Selection

In practice, an approximate solution to the 4D-Var minimization problem (3)-(4) is obtained by performing m-iterations (14) with the initial guess $\mathbf{x}_0^{(0)} = \mathbf{x}_0^b$ and the first-order optimality condition (29) is only approximately satisfied. The analysis $\mathbf{x}_0^a$ is expressed as

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \sum_{i=0}^{m-1} \alpha_i \mathbf{d}^{(i)} \tag{67}$$

such that the analysis increment $\delta\mathbf{x}_0^a$ is an element of the vector space spanned by the descent directions $\delta\mathbf{x}_0^a \in \mathcal{S} = Span\{\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(m-1)}\}$. Denoting $\{\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{m-1}\}$ an orthonormal basis to $\mathcal{S}$, the reduced-space is thus consistent to the 4D-Var minimization process and the approximation (60) is exact. It is noticed that if the conjugate gradient or the BFGS method is implemented and the cost functional is quadratic (for instance, in the incremental 4D-Var formulation) then $\mathbf{x}_0^a$ is the minimizer of J over the set $\{\mathbf{x}_0^b + \mathcal{S}\}$. In this particular case, the first-order optimality condition is satisfied in the reduced space $\nabla_{\boldsymbol{\zeta}} \widehat{J}(\boldsymbol{\zeta}^a) = \mathbf{0}$, although in general $\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^a) \neq \mathbf{0}$.

The reduced basis may be enhanced by appending the vector $\boldsymbol{\psi}_m$ defined as

$$\boldsymbol{\psi} = \nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) - \boldsymbol{\Pi}_{\mathcal{S}} \nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a), \quad \boldsymbol{\psi}_m = \boldsymbol{\psi}/\|\boldsymbol{\psi}\| \tag{68}$$

that is orthogonal to each of the basis vectors $\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{m-1}$, $\boldsymbol{\psi}_m \in \{\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{m-1}\}^{\perp}$. Therefore, $\nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) \in Span\{\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{m-1}, \boldsymbol{\psi}_m\}$ and with this selection of the basis functions the reduced-space provides an exact representation of the forecast sensitivity to analysis $\boldsymbol{\Pi}\nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a) = \nabla_{\mathbf{x}_0} e(\mathbf{x}_0^a)$. From (36) and (66) an error estimate to the reduced-order approximation (64) may be obtained as

$$\boldsymbol{\eta} - \boldsymbol{\Psi}\widehat{\boldsymbol{\eta}} = \left(\mathbf{A}\boldsymbol{\Pi}\mathbf{A}^{-1} - \mathbf{I}_{n\times n}\right) \boldsymbol{\Psi}\widehat{\boldsymbol{\eta}} \tag{69}$$

where $\mathbf{A}$ denotes the inverse of the full Hessian matrix (32) and $\mathbf{I}_{n\times n}$ is the $n \times n$ identity matrix.

NUMERICAL RESULTS

Illustrative numerical results are presented with the finite volume global shallow-water (SW) model of Lin and Rood [48] at a resolution of $2.5^\circ \times 2.5^\circ$ and with a time step $\Delta t = 600$s. The state vector is $\mathbf{x} = (h, u, v)$ where h is the geopotential height and u and v are the zonal and meridional wind velocities, respectively. The first-order adjoint associated with the SW model was developed in the work of Akella and Navon [49], whereas the second-order adjoint was developed by Daescu [31].

An idealized 4D-Var DAS is considered in the twin experiments framework: a reference initial state $\mathbf{x}_0^t$ ("the truth") is taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40 500 hPa data valid at 0600 UTC 15 March 2002; the background estimate $\mathbf{x}_0^b$ to $\mathbf{x}_0^t$ is obtained from a 6-hour integration of the SW model initialized at $t_0 - 6$h with ECMWF ERA-40 500 hPa data valid at 0000 UTC 15 March 2002. At the verification time $t_v = t_0 + 30$h we consider the reference state $\mathbf{x}_v^t = \mathcal{M}_{t_0 \to t_v}(\mathbf{x}_0^t)$ and the forecast $\mathbf{x}_v^f = \mathcal{M}_{t_0 \to t_v}(\mathbf{x}_0^a)$. The forecast error is displayed in Fig. 1 using a total energy norm to obtain grid-point values (units of m^2 s^{-2}).

"Observational data" for the assimilation procedure is generated from a model trajectory initialized with $\mathbf{x}_0^t$ and corrupted with random errors taken from a normal distribution $\mathcal{N}(0, \sigma^2)$. The standard deviation is chosen $\sigma_h = 5$ m for the height and $\sigma_u = \sigma_v = 0.5$ m s^{-1} for the velocities and the observation errors

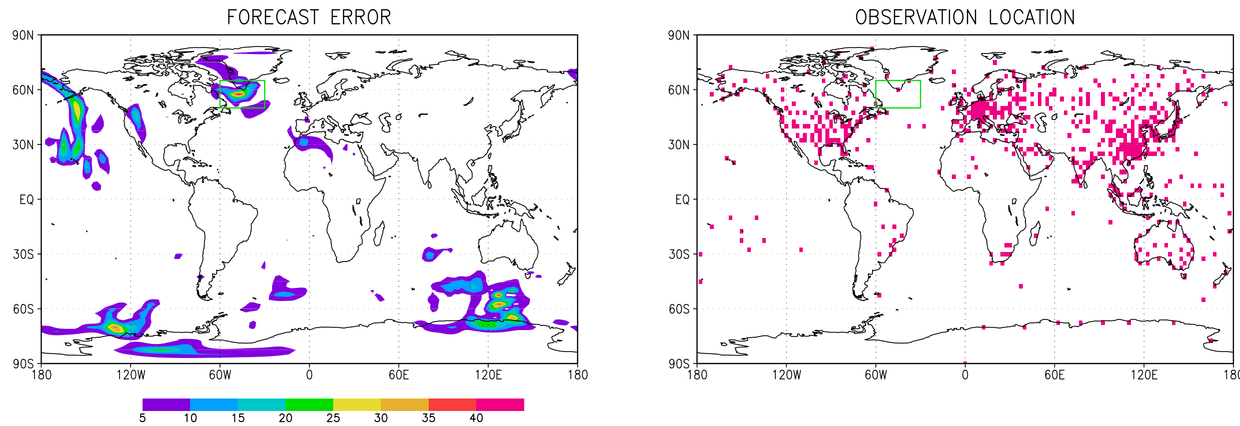

Figure 1. Forecast error ($\mathrm{m}^2\ \mathrm{s}^{-2}$) at $t_v = t_0 + 30\,\mathrm{h}$ and the configuration of the observing system in 4D-Var. The selection of the verification domain at t_v is indicated by the rectangular region.

are assumed to be uncorrelated. The errors in the prior estimate to the initial condition are assumed to be uncorrelated and the standard deviations assigned to the background errors are $\sigma_h^b = 10$ m, $\sigma_u^b = \sigma_v^b = 1$ m s^{-1}. Observational data sets are provided at a time increment of one hour over the data assimilation interval $[t_0, t_0 + 6\mathrm{h}]$ at locations taken from the actual location of the radiosonde observations in a realistic data assimilation system and projected to the nearest model grid point. The observation operator $\mathbf{h}$ is thus a matrix with entries 0 and 1 only. The configuration of the observing system is displayed in Fig. 1. At each observing time there are 572 observations for each of the state variables h, u, and v, whereas the discrete state vector $\mathbf{x}$ is of order $n \sim 3 \times 10^4$.

An increased forecast error is noticed over the region $\mathcal{D}_v = [50^\circ N, 65^\circ N] \times [60^\circ W, 30^\circ W]$, as indicated in Fig. 1. The ability to obtain 4D-Var sensitivity information is illustrated using a forecast error functional defined as the forecast error over $\mathcal{D}_v$ measured in a total energy metric

$$e(\mathbf{x}_0) = (\mathbf{x}_v^f - \mathbf{x}_v^t)^{\mathrm{T}} \mathbf{P}^{\mathrm{T}} \mathbf{E} \mathbf{P} (\mathbf{x}_v^f - \mathbf{x}_v^t) \tag{70}$$

where $\mathbf{P}$ denotes the projection operator on $\mathcal{D}_v$.

Forecast Sensitivity to 4D-Var Input Parameters

A key ingredient to the 4D-Var sensitivity analysis is the solution to the large-scale linear system (36). This computation was performed in the full space using a SOA model to provide the Hessian-vector products required in the CG iteration. The ability of a reduced-order approach to provide sensitivity information is also investigated using a reduced space of dimension $m = 100$ based on the directions $\mathbf{d}^{(i)}$ generated during the limited-memory L-BFGS iteration to the minimization of the 4D-Var cost functional (3). The reduced-order approach entailed significant computational savings and a reduction of as much as 75% in the CPU time as compared with the full space sensitivity estimation.

The u- and v-components of the vector $\boldsymbol{\eta}$ and the corresponding reduced order approximation $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}}$ are displayed in Fig. 2. It is noticed that the reduced order approach is able to closely match the "shape" of the full space solution, however the amplitude of the u- and v-components of $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}}$ is in general lower, as compared with the u- and v-components of $\boldsymbol{\eta}$. To further illustrate this aspect, in Fig. 3 it is shown the configuration of the velocity components of the full state solution $\boldsymbol{\eta}$ and the reduced order solution $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}}$ associated with a forecast error measure defined over the global domain. We recall that in the experiments presented here the background error covariance is specified as a diagonal matrix and $\sigma_u^b = \sigma_v^b = 1$ m. Therefore, from (45) it follows that the u- and v-components of the vector $\boldsymbol{\eta}$ are identified respectively, with the u- and v-components of the forecast sensitivity to the background state $\mathbf{x}_0^b$.

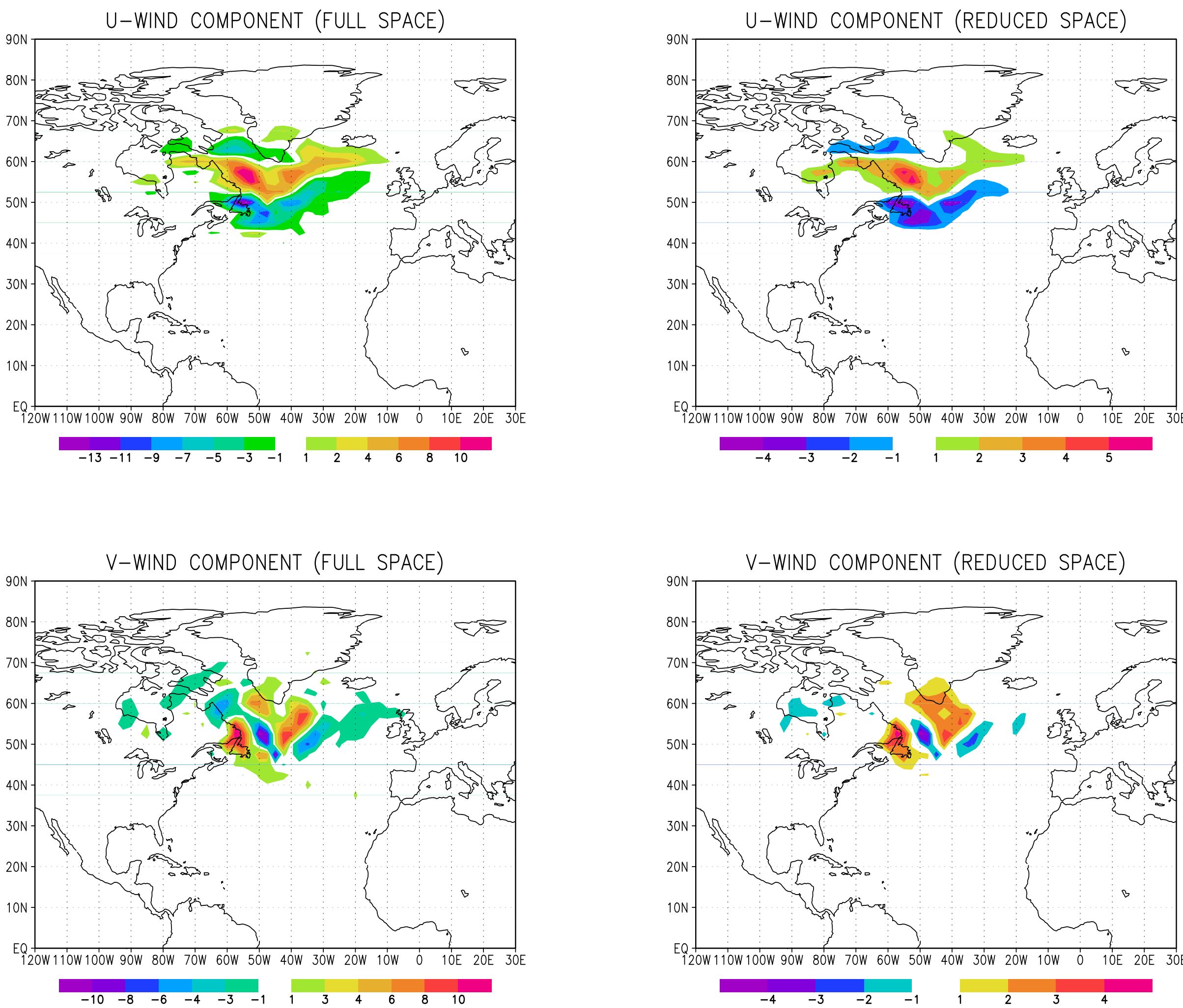

Figure 2. The velocity components of the full state solution $\boldsymbol{\eta}$ and the corresponding reduced order approximation $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}}$. The forecast error measure is defined over the verification domain $[50^{\circ}N, 65^{\circ}N] \times [60^{\circ}W, 30^{\circ}W]$.

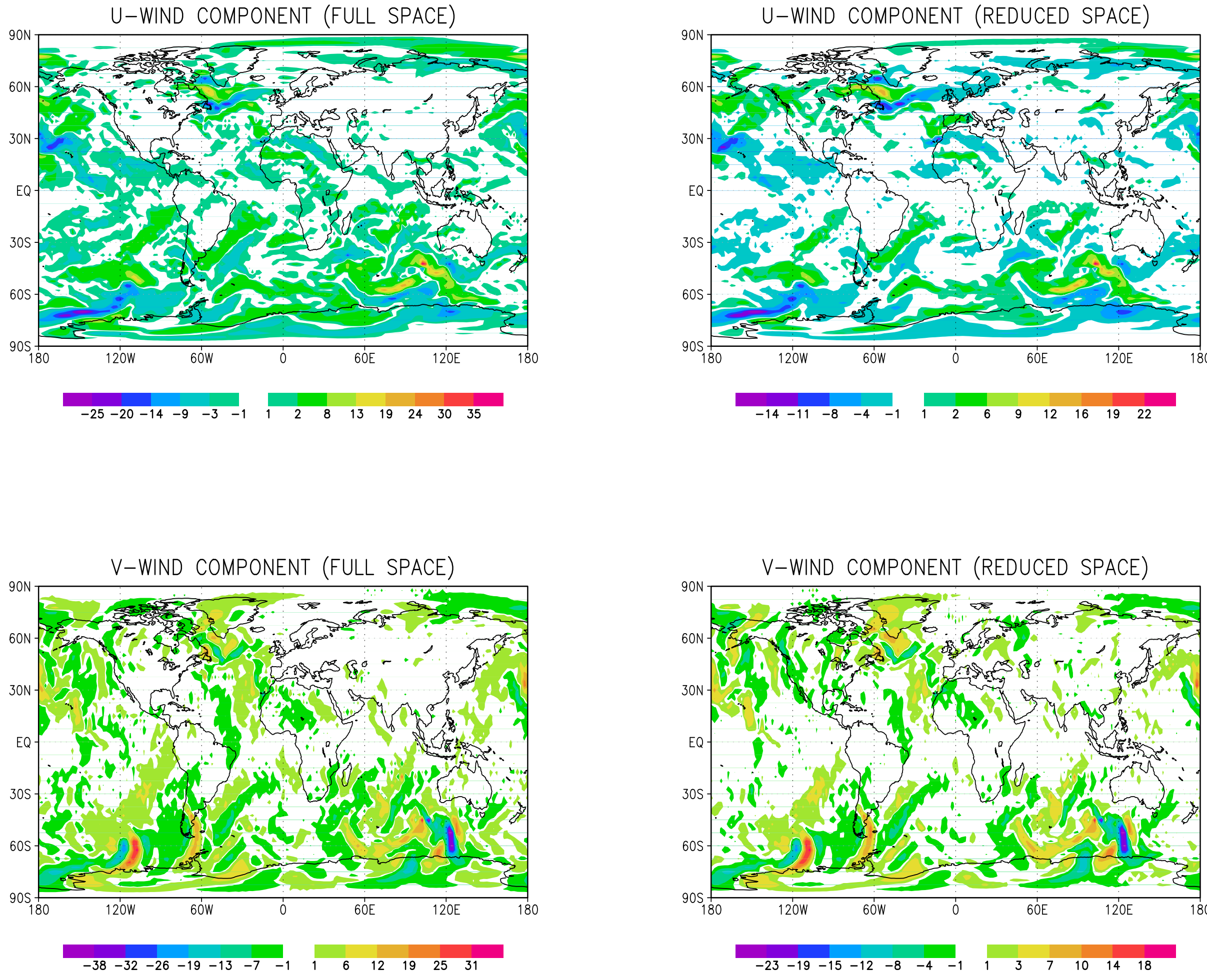

Figure 3. The velocity components of the full state solution $\boldsymbol{\eta}$ and the corresponding reduced order approximation $\boldsymbol{\Psi}\widehat{\boldsymbol{\eta}}$ associated with a forecast error functional defined over the global domain.

The forecast sensitivity to observations and to the specification of the observation error variance are evaluated according to (43) and (44), respectively. As a measure of the forecast sensitivity to observation and error variance at each data location over the assimilation time interval we consider the time cumulative magnitude of the sensitivities $\sum_{i=0}^{N} |\nabla_{\mathbf{y}_i} e(\mathbf{x}_0^a)|$ and $\sum_{i=0}^{N} |\nabla_{\boldsymbol{\sigma}_i^2} e(\mathbf{x}_0^a)|$, respectively. For each of the u, v, and h data components, the location of the observations and error variances of largest forecast sensitivity is displayed in Fig. 4. The sensitivity analysis reveals that the selected forecast error aspect exhibits a large sensitivity with respect to only a few of the observations in the DAS, located in the vicinity of the verification domain. A distinct configuration and magnitude is noticed for each observation component, indicating that observations and error variances whose specifications have a potentially large forecast impact are closely determined by the data type.

A comparison between the observation sensitivity estimates in the full space and the reduced-order estimates indicates that the reduced-order approach is able to properly identify the data locations of largest sensitivity magnitude. Illustrative results are shown in Fig. 5 for the u-wind data at the end of the assimilation window, $t_N = t_0 + 6$ h. These results also indicate that the observation sensitivity estimates in the reduced-order approach have larger magnitudes, as compared to the full state estimates. The reduced-order approach is thus able to properly identify the regions of high background sensitivity and the location

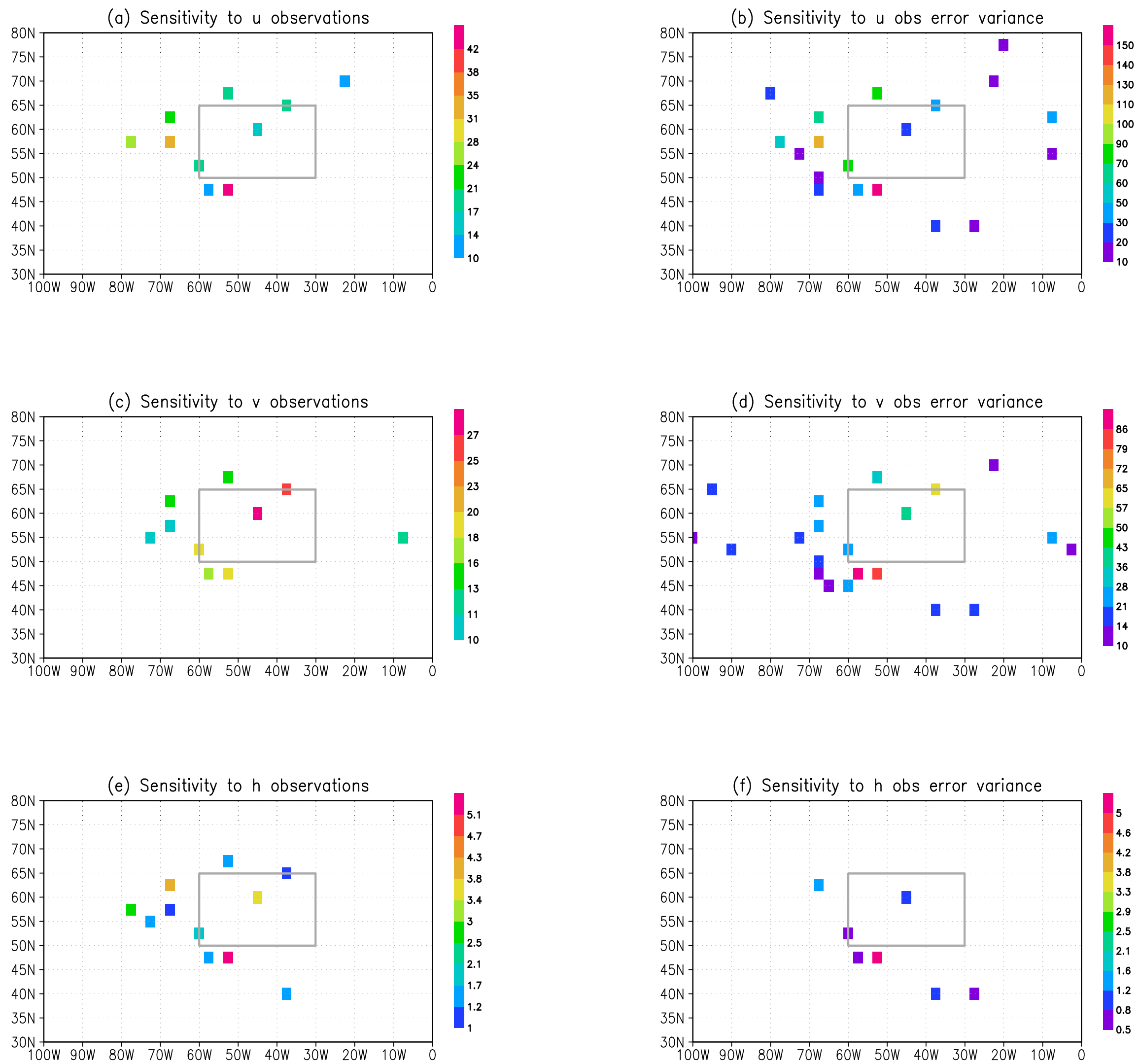

Figure 4. Location of observations with forecast sensitivity of largest magnitude. Time cumulative magnitudes of the observation and error-variance sensitivities are displayed for the h, u, and v observation components.

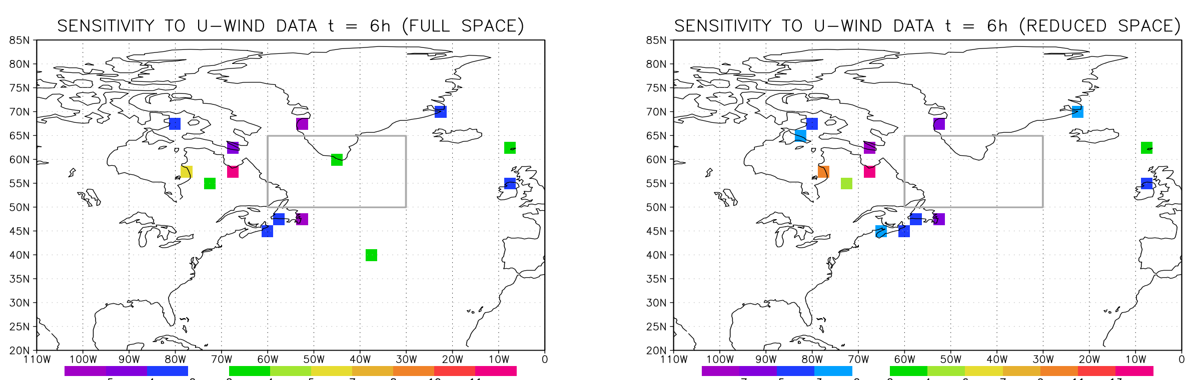

Figure 5. Data locations of largest u-observation sensitivity magnitude at $t_N = t_0+6$ h. Results shown for sensitivity estimates obtained in the full space (left graphic) and in the reduced space (right graphic).

of observations of increased sensitivity; however, it underestimates the background sensitivity values and overestimates the observation sensitivity values, as compared with the full state estimates.

CONCLUSIONS

The combined information derived from the adjoint of the forecast model and the adjoint of the data assimilation system allows the estimation of the model forecast sensitivity with respect to each of the DAS input components. Adjoint-DAS techniques have been mainly considered for a linear analysis scheme and the increased computational burden associated with second-order derivative information is a limiting factor to the implementation in a nonlinear context. This chapter provided a review of the sensitivity equations of a nonlinear 4D-Var DAS and a practical approach to alleviate the computational cost associated with the sensitivity estimation by projection into a low-order subspace. The simplicity of the shallow water model allowed the implementation of a second-order adjoint model associated with the nonlinear 4D-Var formulation and numerical estimation of the sensitivities in both full and reduced-order space. Idealized 4D-Var experiments indicate that the reduced-order approach is able to properly identify those state components and observation locations of increased forecast sensitivity while providing significant computational savings.

The methodologies presented here are general and may find applications to other areas where models and measurements are used to provide improved predictions such as air quality forecasting, climate modeling, oceanography, and fluid dynamics. A rich set of novel applications may be envisaged by performing the sensitivity analysis to the specification of the covariance models used to represent the information error statistics. The weak-constraint 4D-Var formulation allows to account for modeling errors in the analysis scheme and an extension of the sensitivity analysis tools may be envisaged to incorporate the forecast sensitivity and estimation of model error covariance parameters, $\mathbf{x}^a = \mathbf{x}^a(\mathbf{y}, \mathbf{R}, \mathbf{x}_0^b, \mathbf{B}, \mathbf{Q})$. Research in this area is at an incipient stage and further developments are needed in both theoretical and implementation aspects of sensitivity analysis, parameter estimation, forecast impact assessment, and uncertainty quantification in nonlinear data assimilation.

ACKNOWLEDGEMENT

The work of D. N. Daescu was supported by the National Science Foundation under grant DMS-0914937. Prof. I.M. Navon acknowledges the support of National Science Foundation under CMG grant ATM-0931198.

REFERENCES

[1] Jazwinski AH. *Stochastic Processes and Filtering Theory.* Academic Press 1970; 376pp.

[2] Daley R. *Atmospheric Data Analysis.* Cambridge University Press 1991; 457pp.

[3] Kalnay E. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press 2002; 341pp.

[4] Blum J, Le Dimet FX, Navon IM. Data assimilation for geophysical fluids. In Ciarlet PG, Editor(s), Handbook of Numerical Analysis, Elsevier, 2009; 14: 385–441.

[5] Thépaut JN, Andersson E. The global observing system. In Lahoz W, Khattatov B, Ménard R (eds.) *Data Assimilation: Making Sense of Observations.* Springer 2010; 263–281.

[6] Lorenc AC. Modelling of error covariances by 4D-Var data assimilation. Q J R Meteorol Soc 2003; 129: 3167–3182.

[7] Hamill TM, Snyder C. Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. Mon Weather Rev 2002; 130: 1552–1572.

[8] Janjić T, Cohn SE. Treatment of observation error due to unresolved scales in atmospheric data assimilation. Mon Weather Rev 2006; 134: 2900–2915.

[9] Frehlich R. The definition of 'truth' for Numerical Weather Prediction error statistics. Q J R Meteorol Soc 2011; 137: 84–98.

[10] Purser RJ, Parrish D, Masutani M. Meterorological observational data compression; an alternative to conventional "super-obbing" 2000; Office Note 430, National Centers for Environmental Prediction, Camp Springs, MD.

[11] Atlas R. Atmospheric observations and experiments to assess their usefulness in data assimilation. J Meteorol Soc Japan 1997; 75 (1B): 111–130.

[12] Bouttier F, Kelly G. Observing-system experiments in the ECMWF 4D-Var data assimilation system. Q J R Meteorol Soc 2001; 127: 1469–1488.

[13] Langland RH, Baker NL. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus 2004; 56A: 189–201.

[14] Kelly G, Thépaut JN, Buizza R, Cardinali C. The value of observations. I: Data denial experiments for the Atlantic and the Pacific. Q J R Meteorol Soc 2007; 133: 1803–1815.

[15] Cardinali C. Monitoring the observation impact on the short-range forecast. Q J R Meteorol Soc 2009; 135: 239–250.

[16] Gelaro R, Langland RH, Pellerin S, Todling R. The THORPEX observation impact intercomparison experiment. Mon Weather Rev 2010; 138: 4009–4025.

[17] Bennett AF. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press 2002; Cambridge, 234 pp.

[18] Rabier F. Overview of global data assimilation developments in numerical weather-prediction centres. Q J R Meteorol Soc 2005; 131: 3215–3233.

[19] Liu DC, Nocedal J. On the limited memory BFGS method for large-scale optimization. Math Program 1989; 45: 503–528.

[20] Bertsekas DP. *Nonlinear Programming*. Athena Scientific 1995; 646 pp.

[21] Nocedal J, Wright SJ. *Numerical Optimization*. Springer-Verlag, New York, 2000; 636pp.

[22] Courtier P, Thépaut JN, Hollingsworth A. A strategy of operational implementation of 4D-Var using an incremental approach. Q J R Meteorol Soc 1994; 120: 1367–1388.

[23] Trémolet Y. Diagnostics of linear and incremental approximations in 4D-Var. Q J R Meteorol Soc 2004; 130: 2233–2251.

[24] Trémolet Y. Incremental 4D-Var convergence study. Tellus A 2007; 59: 706–718.

[25] Lawless AS, Gratton S, Nichols NK. An investigation of incremental 4D-Var using non-tangent linear models. Q J R Meteorol Soc 2005; 131: 459–476.

[26] Cacuci DG. *Sensitivity & Uncertainty Analysis, Volume 1: Theory*. Chapman and Hall/CRC 2003; 304pp.

[27] Cacuci DG, Ionescu-Bujor M, Navon M. *Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems*. CRC Press 2005; 368pp.

[28] *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*. CRC Press 1996; 275pp.

[29] Saltelli A, Chan K, Scott EM. *Sensitivity Analysis*. Wiley 2009; 492pp.

[30] Le Dimet FX, Ngodock HE, Luong B, Verron J. Sensitivity analysis in variational data assimilation. J Meteorol Soc Japan 1997; 75: 245–255.

[31] Daescu DN. On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. Mon Weather Rev 2008; 136: 3050–3065.

[32] Le Dimet FX, Navon IM, Daescu DN. Second-order information in data assimilation. Mon Weather Rev 2002; 130: 629–648.

[33] Giering R, Kaminski T. Recipes for adjoint code construction. ACM Trans Math Software 1998; 24: 437–474.

[34] Griewank A, Walther A. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* 2^{nd} Edition, SIAM 2008; 438pp.

[35] Nash SG, Nocedal J. A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization. SIAM J Optim 1991; 1: 358–372.

[36] Schlick T, Fogelson A. TNPACK - a truncated Newton minimization package for large-scale problems: I. Algorithm and usage. ACM Trans Math Software 1992; 18: 46–70.

[37] Morales JL, Nocedal J. Enriched methods for large-scale unconstrained optimization. Comput Optim Appl 2002; 21: 143–154.

[38] Daescu DN, Navon IM. An analysis of a hybrid optimization method for variational data assimilation. Int J Comp Fld Dyn 2003; 17: 299–306.

[39] Observation and background adjoint sensitivity in the adaptive observation-targeting problem. Q J R Meteorol Soc 2000; 126: 1431–1454.

[40] Daescu DN, Todling R. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. Q J R Meteorol Soc 2010; 136: 2000–2012.

[41] Daescu DN, Langland RH. Error covariance sensitivity and impact estimation with adjoint 4D-Var: theoretical aspects and first applications to NAVDAS-AR. Q J R Meteorol Soc 2012; DOI: 10.1002/qj.1943 (in press).

[42] Trémolet Y. Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus* **A** 2008; 60: 964–978.

[43] Zhu Y, Gelaro R. Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. Mon Weather Rev 2008; 136: 335–351.

[44] Efficiency of a POD-based reduced second-order adjoint model in 4D-Var data assimilation. Int J Num Meth Fld 2007; 53: 985–1004.

[45] Daescu DN, Navon IM. A dual-weighted approach to order reduction in 4DVAR data assimilation. Mon Weather Rev 2008; 136: 1026–1041.

[46] Fisher M, Courtier P. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. European Centre for Medium-Range Weather Forecasts Tech Mem 220, 1995; 28pp. Reading, UK.

[47] Leutbecher M. A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network. J Atmos Sci 2003; 60: 729–742.

[48] Lin SJ, Rood RB. An explicit flux-form semi-Lagrangian shallow-water model on the sphere. Q J R Meteorol Soc 1997; 123: 2477–2498.

[49] Akella S, Navon IM. A comparative study of the performance of high-resolution advection schemes in the context of data assimilation. Int J Numer Methods Fluids 2006; 51: 719–748.