Tellus
SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2008), 60A, 964–978

TELLUS

Computation of observation sensitivity and observation impact in incremental variational data assimilation

By YANNICK TRÉMOLET*, *Global Modeling and Assimilation Office, Code 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*

(Manuscript received 14 January 2008; in final form 20 May 2008)

ABSTRACT

We discuss the computation of observation sensitivities and observation impact for incremental variational data assimilation (VDA), accounting for the inner and outer loops. To fully account for the outer loops, a second-order adjoint of the data assimilation system is required, which makes it impractical for an operational data assimilation system. However, some approximations can be made that allow useful results to be obtained with multiple outer loop iterations, in particular, for observation impact studies.

Two algorithms are presented to compute the adjoint of the inner loop minimization, and their merits are discussed. Validation results are given for both of these algorithms. We show that one algorithm, based on the adjoint of an approximation of the inverse of the Hessian of the cost function, can also be used to investigate some convergence aspects of the incremental VDA inner loop. Because it is computationally inexpensive, the proposed algorithm could be used to monitor an operational system routinely. We give some numerical results illustrating the impact of observations in successive outer loop iterations.

1. Introduction

As the number of observations being assimilated to produce meteorological forecasts has grown almost exponentially over the years, the data assimilation process has become more and more complex and expensive. It is important to evaluate the impact of all these observations, to assess the cost effectiveness of collecting and assimilating them, and to assess the ability of the data assimilation system to use these observations effectively. One technique to evaluate the value of observations is by way of observing system experiments (OSE), but they tend to be very expensive because they can be performed for only one subset of observations at a time. Ensemble techniques have also been applied to assess the potential impact of future observing system, for example by Tan et al. (2007). This technique also has the disadvantage of cost since an ensemble of data assimilation systems is required. Both techniques have the disadvantage that modifying the observation system can change the value of remaining observations. For these reasons, these techniques cannot be used routinely in operational systems.

Recently, Baker and Daley (2000) have shown that observation sensitivity can be computed using the adjoint of the data assimilation system. The goal of observation sensitivity studies is to evaluate the sensitivity of a measure of an aspect of a forecast to the observations that was used to produce this forecast, through the data assimilation process. This is often a measure of forecast quality, but the results apply to any measurable aspect of the forecast.

More precisely, we are interested here in a scalar measure $F(\mathbf{x}_f)$ of an aspect of a forecast. The forecast $\mathbf{x}_f$ is issued from an analysis $\mathbf{x}_a$ using an atmospheric model $\mathcal{M}$:

$$\mathbf{x}_f = \mathcal{M}(\mathbf{x}_a).$$

The analysis is itself the result of a data assimilation process, represented by the operator G:

$$\mathbf{x}_a = G(\mathbf{x}_b, \mathbf{y}).$$

where $\mathbf{x}_b$ represents the background state and $\mathbf{y}$ represents the observations. We define the measure of the forecast aspect $\mathcal{F}$ as a function of the inputs to the data assimilation system $\mathbf{x}_b$ and $\mathbf{y}$ by:

$$\mathcal{F}(\mathbf{x}_b, \mathbf{y}) = F \circ \mathcal{M} \circ G(\mathbf{x}_b, \mathbf{y}),$$

where $\circ$ denotes the composition of functions. Applying the usual derivation chain rule and the definition of the adjoint and gradient of the scalar function $\mathcal{F}$, the sensitivity of $\mathcal{F}$ to the observations is:

$$\frac{\partial \mathcal{F}}{\partial \mathbf{y}} = \mathbf{G}^T \mathbf{M}^T \frac{\partial F}{\partial \mathbf{x}_f}. \quad (1)$$

*Correspondence.
e-mail: ytremolet@gmao.gsfc.nasa.gov
DOI: 10.1111/j.1600-0870.2008.00349.x

Another application for the adjoint of the data assimilation system is the computation of observation impact. The impact of an observation on the measure $\mathcal{F}$ can be defined as the variation of $\mathcal{F}$ that is due to the increment generated by that observation. This impact can be computed using Taylor series expansions of various order. There is a very simple relationship between observation sensitivity and observation impact in a linear data assimilation system. In a variational data assimilation (VDA) system with multiple outer loops, there is no direct relationship between observation sensitivity and observation impact, but the adjoint of the data assimilation system can still be used to compute observation impact as will be shown in Section 3.4.

The adjoint model $\mathbf{M}^T$ is well known in meteorology and used extensively in 4D-Var and other sensitivity applications. Measures that depend directly on the analysis values can also be used. In that case, the forecast model is not present in the expression above, and the sensitivity of the analysis to the observations is measured. The adjoint of the data assimilation system $\mathbf{G}^T$ is more complex, and its properties and use have only recently started to be investigated.

In observation space data assimilation systems, the adjoint of the data assimilation system can be computed indirectly, as was done by Langland and Baker (2004), Xu et al. (2006) or Pellerin et al. (2006). In model space data assimilation systems, another approach comprises developing the adjoint of the data assimilation system, as experimented by Zhu and Gelaro (2008), with the grid-point statistical interpolation system (GSI, Wu et al., 2002), which is a 3D-Var model space data assimilation system. However, recent development of a 4D-Var version of the GSI made this adjoint obsolete. It would be a difficult and expansive work to re-write the adjoint of the 4D-Var system. We are presenting here another approach for computing the adjoint of a model space VDA system.

The objective of this paper is to propose an algorithm to compute the sensitivity to observations in model space incremental VDA systems. We present methods for calculating the data assimilation adjoint indirectly and propose an algorithm which is computationally very inexpensive. Moreover, most algorithms investigated so far have assumed the data assimilation system to be linear, even though it is known that in VDA, in 4D-Var in particular, non-linear effects, accounted for through the outer loop iterations, are important to improve the analysis and forecast quality. Therefore, we will also investigate potential approaches to account for non-linear outer loops and highlight the approximations that make observation sensitivity computations possible.

The paper is organized as follows: in the next section, we introduce the incremental VDA algorithm, based on outer and inner loops. In Section 3 we formulate the adjoint of the outer loop iteration. We show that the full adjoint of the outer loops iterations would be very complex and give some possible approximations. The computation of observation impact with multiple outer loop iterations is also discussed. In Section 4, two algorithms are described to compute the adjoint of the inner loop minimization of the incremental data assimilation system. In Section 5 we give some numerical results, validating the adjoint algorithms and showing how observations are used in successive inner and outer iterations, before the conclusion.

2. Incremental variational data assimilation

In this paper, we focus on model space VDA systems such as 3D-Var or 4D-Var. In these systems, the analysis is the result of the minimization of a cost function:

$$J(\mathbf{x})=\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+\frac{1}{2}(\mathcal{H}(\mathbf{x})-\mathbf{y})^T\mathbf{R}^{-1}(\mathcal{H}(\mathbf{x})-\mathbf{y}),$$

where $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices and $\mathcal{H}$ is the observation operator that computes the observation estimates corresponding to the atmospheric state $\mathbf{x}$. In 3D-Var, the control variable and observation operator are three-dimensional. In 4D-Var, the observation operator is four-dimensional and we assume that the forecast model is embedded in the observation operator. In weak constraint 4D-Var, the control vector and observation operator are both four-dimensional. More details on the various formulations of VDA in model space are given, for example, by Trémolet (2006). The results given here are general and apply to any model space VDA system.

In practice, the cost function defined above is non-linear and difficult to minimize. An iterative Gauss-Newton approach is used: the increment is chosen as the control variable of the problem and the observation operator is linearized around the current state estimate. This defines an approximate quadratic cost function which is minimized. The process is then repeated until a satisfactory solution has been found: this is the outer loop of 3D-Var or 4D-Var. The inner loop designates the iterations of the minimization algorithm used to minimize the quadratic cost function within each outer loop iteration. This algorithm is known as incremental VDA and was introduced by Courtier et al. (1994).

At iteration j of the outer loop, the quadratic cost function being minimized is:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x}-\mathbf{b}_j)^T\mathbf{B}^{-1}(\delta\mathbf{x}-\mathbf{b}_j) + \frac{1}{2}(\mathbf{H}_j\delta\mathbf{x}-\mathbf{d}_j)^T\mathbf{R}^{-1}(\mathbf{H}_j\delta\mathbf{x}-\mathbf{d}_j), \quad (2)$$

where $\delta\mathbf{x}=\mathbf{x}-\mathbf{x}_{j-1}$, $\mathbf{b}_j=\mathbf{x}_b-\mathbf{x}_{j-1}$, $\mathbf{d}_j=\mathbf{y}-\mathcal{H}(\mathbf{x}_{j-1})$ and $\mathbf{H}_j$ is the observation operator linearized around the state estimate $\mathbf{x}_{j-1}$ resulting from the previous iteration. This minimization problem is equivalent to solving the linear system $\nabla J=0$ and the solution of the assimilation problem can be written explicitly. We define the matrix:

$$\mathbf{K}_j=(\mathbf{B}^{-1}+\mathbf{H}_j^T\mathbf{R}^{-1}\mathbf{H}_j)^{-1}\mathbf{H}_j^T\mathbf{R}^{-1}.$$

The solution at outer loop j is:

$$\delta\mathbf{x}_j=\mathbf{K}_j\mathbf{d}_j+(\mathbf{I}-\mathbf{K}_j\mathbf{H}_j)\mathbf{b}_j. \quad (3)$$

The starting point for the minimization is usually the background state so we assume here that $\mathbf{x}_0 = \mathbf{x}_b$. The analysis is $\mathbf{x}_a = \mathbf{x}_m$ where m is the number of outer loop iterations.

If the model and observation operators used in the definition of the cost function are linear, the un-approximated cost function and the data assimilation problem become quadratic. In that case, the last term on the right hand side of eq. (3) vanishes, and the VDA problem is also equivalent to the Kalman filter, and the associated matrix $\mathbf{K}$ is called the Kalman gain matrix. We will generalize the use of the term Kalman gain matrix to $\mathbf{K}_j$ even though it is associated to an intermediate problem, not to the full data assimilation system.

In practice the non-linear observation operator $\mathcal{H}$ might depend on j if quality control decisions are made during the minimization process, and the set of observations being used in each outer loop or the observation error variances varies. For simplicity, we assume in the following that $\mathcal{H}$ is independent of j.

In most operational data assimilation systems, an additional term J_c is added to the cost function to enforce balance in the analysis. J_c can take the form of a normal mode initialization term or of the digital filter initialization term for example. Constraints to the same effect can also be included in the background term of the cost function J_b. Thus we will omit this term or assume that it is embedded in the J_b term.

3. Observation sensitivity with multiple outer loops

3.1. Constant operators

As explained above, the VDA cost function is minimized using an iterative algorithm which can be represented at iteration j by an operator G_j such that

$$(\mathbf{x}_j, \mathbf{x}_b, \mathbf{y}) = G_j(\mathbf{x}_{j-1}, \mathbf{x}_b, \mathbf{y}).$$

The full data assimilation system can then by represented by the operator $G = G_m \circ \cdots \circ G_1$. Since $\mathbf{x}_b$ and $\mathbf{y}$ do not vary, they could be omitted from the definition of G_j. They are kept here explicitly as a reminder that they are the inputs to the data assimilation system.

For this section, the operators G_j are assumed constant relative to the variables of the problem. This assumption means that, at a given iteration, although the previous state estimate might be used to define the operator G_j (often as a trajectory for linearization), G_j will be considered a constant for all differentiation operations. In other words, the terms of the form $\partial G_j/\partial \mathbf{x}_{j-1}$ are omitted from the derivations. From eq. (3), which is a linear expression, the linearized operator corresponding to the minimization is obtained immediately. However, one iteration comprises two steps: the computation of the departures in the outer loop and the minimization itself. Iteration j is written as

$$\begin{aligned}
\mathbf{d}_j &= \mathbf{y} - \mathcal{H}(\mathbf{x}_{j-1}),\\
\delta\mathbf{x}_j &= \mathbf{K}_j\mathbf{d}_j + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)(\mathbf{x}_b - \mathbf{x}_{j-1}),\\
\mathbf{x}_j &= \mathbf{x}_{j-1} + \delta\mathbf{x}_j.
\end{aligned}$$

To compute the sensitivities, the tangent linear iteration is derived first:

$$\begin{aligned}
\tilde{\mathbf{d}}_j &= \tilde{\mathbf{y}} - \mathbf{H}_j\tilde{\mathbf{x}}_{j-1},\\
\delta\tilde{\mathbf{x}}_j &= \mathbf{K}_j\tilde{\mathbf{d}}_j + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)(\tilde{\mathbf{x}}_b - \tilde{\mathbf{x}}_{j-1}),\\
\tilde{\mathbf{x}}_j &= \tilde{\mathbf{x}}_{j-1} + \delta\tilde{\mathbf{x}}_j,
\end{aligned}$$

where the $\tilde{(.)}$ notation represents the tangent linear variables. Substituting $\delta\tilde{\mathbf{x}}_j$ and $\tilde{\mathbf{d}}_j$ in the last equation, the linearized operator corresponding to G_j is

$$\tilde{\mathbf{x}}_j = \mathbf{K}_j\tilde{\mathbf{y}} + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\tilde{\mathbf{x}}_b.$$

Note that $\tilde{\mathbf{x}}_{j-1}$ does not appear in this expression. This is a consequence of the fact that when the problem is linear, the solution is unique and independent of the starting point of the minimization. The adjoint of this operator for iteration j is

$$\begin{aligned}
\mathbf{y}^* &= \mathbf{y}^* + \mathbf{K}_j^T\mathbf{x}_j^*,\\
\mathbf{x}_b^* &= \mathbf{x}_b^* + (\mathbf{I} - \mathbf{H}_j^T\mathbf{K}_j^T)\mathbf{x}_j^*,\\
\mathbf{x}_{j-1}^* &= 0,
\end{aligned}$$

where the notation $(.)^*$ represents the adjoint variables. The operator $\mathbf{K}_j^T$ needs to be applied once per outer iteration. As described in Section 4, this can be done easily within the VDA algorithm. The other operator appearing in this algorithm, $\mathbf{H}_j^T$, is already available in the VDA algorithm. The adjoint of the full VDA system can thus be obtained with a very limited amount of code development, at least if the operators involved are assumed state independent.

Applying the usual chain rule for the adjoint, the adjoint outer iterations are performed backwards, starting with $\mathbf{y}^* = 0$, $\mathbf{x}_b^* = 0$ and $\mathbf{x}_m^* = \partial\mathcal{F}/\partial\mathbf{x}_a$. However, because there is no sensitivity to $\mathbf{x}_{m-1}$, the last line of the adjoint of outer iteration m is $\mathbf{x}_{m-1}^* = 0$. That implies that the contributions from the adjoints of all other iterations are zero. Thus, under the assumptions made in this section, the computation of the sensitivity to observations only requires the adjoint of the last outer iteration and yields

$$\frac{\partial\mathcal{F}}{\partial\mathbf{y}} = \mathbf{K}_m^T\frac{\partial\mathcal{F}}{\partial\mathbf{x}_a} \quad\text{and}\quad \frac{\partial\mathcal{F}}{\partial\mathbf{x}_b} = (\mathbf{I} - \mathbf{H}_m^T\mathbf{K}_m^T)\frac{\partial\mathcal{F}}{\partial\mathbf{x}_a}.$$

The intermediate state estimates are only a tool to get to the solution but have no meaning as such. Once the quadratic problem being solved in the last outer iteration has been defined, its solution does not depend on the first guess, that is, $\mathbf{x}_{m-1}$, but only on the observations and background that define the problem. One could choose to use $\mathbf{x}_b$ or any other value as the first guess for all minimizations and obtain the same analysis. Since using the result of the previous minimization as the first guess reduces the computational cost without changing the result, it is a better choice to start the minimization. The fact that, with the assumptions made here, the only role of $\mathbf{x}_{m-1}$ is as a first guess, explains that there is no sensitivity to it or earlier intermediate estimates.

In most cases, sensitivity computations are performed assuming a single outer loop, often the first one. This first result shows that, in fact, it is better to use the adjoint of the last outer loop (This does not involve any additional cost or complexity). The difference is in the trajectory used to define $\mathbf{K}_j^T$ and $\mathbf{H}_j^T$, which is more accurate in the last iteration. The consequence is that $\mathbf{K}_m^T$ and $\mathbf{H}_m^T$ are more representative of the derivatives at the analysis point whereas, in the first iteration, $\mathbf{K}_1^T$ and $\mathbf{H}_1^T$ are more representative of the derivatives in the vicinity of the background.

In this section, strong simplifying assumptions have been made to present the observation sensitivity problem in its simplest implementation. In the following two sections, we try to remove these assumptions to compute more accurate sensitivities.

3.2. Accounting for the outer loop

In reality, the increment from the previous outer iteration has an impact on the current outer loop iteration since it is used in the computation of $\mathbf{d}_j$ and $\mathbf{b}_j$. In operational systems, a simplified linear observation operator is often used in the inner loop. For example, a common implementation of 3D-Var is the 3D-FGAT (first guess at appropriate time), where the outer loop accounts for the time evolution, through the use of a forecast model, whereas the inner loop ignores the time dimension. In most 4D-Var systems, simplifications include running the inner loop at lower resolution than the outer loop and, possibly, simpler physics in the tangent linear model. Let $\mathbf{H}_j^{\mathrm{lr}}$ be the simplified observation operator used in the inner loop at iteration j. For clarity, we explicitly use $\mathbf{H}^{\mathrm{hr}}$ to represent the linear observation operator, corresponding to the non-linear observation operator used in the (higher-resolution) outer loop. We define $\mathbf{P}_j$ as the interpolation from the resolution of the outer loops to the resolution of inner loop j. We assume $\mathbf{P}_j$ is linear and that a pseudo-inverse $\mathbf{P}_j^{-1}$ exists. In some applications, such as 3D-FGAT, inner and outer loops might be run at the same resolution. In that case $\mathbf{P}_j$ is the identity. With these notations, the tangent linear iteration becomes

$$
\begin{aligned}
\tilde{\mathbf{d}}_j &= \tilde{\mathbf{y}} - \mathbf{H}_j^{\mathrm{hr}} \tilde{\mathbf{x}}_{j-1}, \\
\delta\tilde{\mathbf{x}}_j &= \mathbf{K}_j \tilde{\mathbf{d}}_j + (\mathbf{I} - \mathbf{K}_j \mathbf{H}_j^{\mathrm{lr}}) \mathbf{P}_j (\tilde{\mathbf{x}}_{\mathrm{b}} - \tilde{\mathbf{x}}_{j-1}), \\
\tilde{\mathbf{x}}_j &= \tilde{\mathbf{x}}_{j-1} + \mathbf{P}_j^{-1} \delta\tilde{\mathbf{x}}_j,
\end{aligned}
$$

where the state estimate $\tilde{\mathbf{x}}_j$ is at outer loop resolution and the increment $\delta\tilde{\mathbf{x}}_j$ is at inner loop resolution. In this system, it is not possible to eliminate $\tilde{\mathbf{x}}_{j-1}$ from the tangent linear iteration: the solution at a given outer loop j is not independent from the first guess. Some linear algebra shows that the adjoint of iteration j is

$$
\begin{aligned}
\mathbf{y}^* &= \mathbf{y}^* + \mathbf{K}_j^T \mathbf{P}_j^{-T} \mathbf{x}_j^*, \\
\mathbf{x}_{\mathrm{b}}^* &= \mathbf{x}_{\mathrm{b}}^* + \mathbf{P}_j^T \left[\mathbf{I} - (\mathbf{H}_j^{\mathrm{lr}})^T \mathbf{K}_j^T\right] \mathbf{P}_j^{-T} \mathbf{x}_j^*, \\
\mathbf{x}_{j-1}^* &= \mathbf{x}_{j-1}^* + \left[\mathbf{P}_j^T (\mathbf{H}_j^{\mathrm{lr}})^T - (\mathbf{H}_j^{\mathrm{hr}})^T\right] \mathbf{K}_j^T \mathbf{P}_j^{-T} \mathbf{x}_j^*.
\end{aligned}
$$

Note that, contrary to the usual definition in 4D-Var, the adjoint variable $\mathbf{x}_j^*$ is defined at the outer loop (high) resolution. This result is similar to the result given in the previous section, except for the additional interpolations between resolutions and, more importantly, the fact that $\mathbf{x}_{j-1}^*$ is non-zero. The correction to $\mathbf{x}_{j-1}^*$ comes from the difference between the inner and outer loops. When the inner and outer loops use the same observation operator, $\mathbf{H}_j^{\mathrm{hr}}$ and $\mathbf{H}_j^{\mathrm{lr}}$ are the same, and the additional term vanishes as described in Section 3.1. The correction term measures the sensitivity to the error introduced by reducing the resolution in the inner loop and possibly other simplifications such as the omission of some of the linear physics in the minimizations. In general, when the inner and outer loop agree well, this term will be small, and the adjoint of the outer loops other than the last one will have a small contribution to the overall sensitivity. Under the assumptions made in this section, the adjoint of all outer iterations is required. The operator $\mathbf{K}_j^T$ is applied once per outer iteration, using the algorithms described in Section 4.

3.3. State dependent operators

Reality is in fact even more complex. The operators that define the inner loop cost function, depend on the current state estimate, which is used as a linearization state. In that case, $\mathbf{H}_j$ and $\mathbf{K}_j$ depend on the trajectory, that is, on $\mathbf{x}_{j-1}$, which itself depends on the observations and background.

As explained in Section 2, the increment at outer loop j is

$$
\delta\mathbf{x}_j = \mathbf{K}_j \mathbf{d}_j + (\mathbf{I} - \mathbf{K}_j \mathbf{H}_j) \mathbf{b}_j. \tag{4}
$$

To simplify the presentation, we focus here on the first term of this expression. We write the matrix product explicitly for the nth component $\delta x_{j,n}$ of $\delta\mathbf{x}_j$:

$$
\delta x_{j,n} = \sum_p k_{j,n,p} d_{j,p},
$$

where $d_{j,p}$ is the pth component of the observation space vector $\mathbf{d}_j$ and the $k_{j,n,p}$ are the coefficients of the matrix $\mathbf{K}_j$(n and m represent indices in control space, p represents indices in observation space). The tangent linear expression for each component of $\delta\mathbf{x}_j$ is

$$
\delta\tilde{x}_{j,n} = \sum_p k_{j,n,p} \tilde{d}_{j,p} + \sum_p \left(\sum_m \frac{\partial k_{j,n,p}}{\partial x_{j-1,m}} \tilde{x}_{j-1,m} \right) d_{j,p}. \tag{5}
$$

Re-arranging the sums gives

$$
\delta\tilde{x}_{j,n} = \sum_p k_{j,n,p} \tilde{d}_{j,p} + \sum_m \left(\sum_p \frac{\partial k_{j,n,p}}{\partial x_{j-1,m}} d_{j,p} \right) \tilde{x}_{j-1,m}.
$$

Let $\boldsymbol{\delta}\mathbf{K}_j$ be the tensor whose coefficients are

$$\hat{k}_{j,n,m,p} = \frac{\partial k_{j,n,p}}{\partial x_{j-1,m}},$$

the product $\boldsymbol{\delta}\mathbf{K}_j \bullet \mathbf{d}_j$ is the matrix whose coefficients are

$$\hat{k}_{j,n,m} = \sum_p \hat{k}_{j,n,m,p} d_{j,p}.$$

The tangent linear expression corresponding to the first term in eq. (4) is

$$\delta\tilde{\mathbf{x}}_j = \mathbf{K}_j\tilde{\mathbf{d}}_j + (\boldsymbol{\delta}\mathbf{K}_j \bullet \mathbf{d}_j)\tilde{\mathbf{x}}_{j-1}.$$

Defining $\boldsymbol{\delta}\mathbf{H}_j$ in a similar way, the second term in eq. (4) can be differentiated. The complete tangent linear expression is

$$\begin{aligned}\delta\tilde{\mathbf{x}}_j = {} & \mathbf{K}_j\tilde{\mathbf{d}}_j + (\boldsymbol{\delta}\mathbf{K}_j \bullet \mathbf{d}_j)\tilde{\mathbf{x}}_{j-1} + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\tilde{\mathbf{b}}_j \\ & -[\boldsymbol{\delta}\mathbf{K}_j \bullet (\mathbf{H}_j\mathbf{b}_j) + \mathbf{K}_j(\boldsymbol{\delta}\mathbf{H}_j \bullet \mathbf{b}_j)]\tilde{\mathbf{x}}_{j-1}.\end{aligned}$$

The adjoint system can thus be written

$$\begin{aligned}\mathbf{y}^* &= \mathbf{y}^* + \mathbf{K}_j^T\mathbf{x}_j^*, \\ \mathbf{x}_b^* &= \mathbf{x}_b^* + (\mathbf{I} - \mathbf{H}_j^T\mathbf{K}_j^T)\mathbf{x}_j^*, \\ \mathbf{x}_{j-1}^* &= \mathbf{x}_{j-1}^* + \big[(\boldsymbol{\delta}\mathbf{K}_j \bullet \mathbf{d}_j)^T - (\boldsymbol{\delta}\mathbf{K}_j \bullet (\mathbf{H}_j\mathbf{b}_j))^T \\ &\quad -(\boldsymbol{\delta}\mathbf{H}_j \bullet \mathbf{b}_j)^T\mathbf{K}_j^T\big]\,\mathbf{x}_j^*.\end{aligned}$$

The terms resulting from the differences and interpolations between the inner and outer loops (Section 3.2) have been omitted here for clarity, they should be added to obtain the full adjoint. The $\boldsymbol{\delta}$ notation denotes the derivatives with respect to the state estimate (or linearization state or trajectory). $\boldsymbol{\delta}\mathbf{K}_j$ is a generalization of the second-order adjoint described by Wang et al. (1992) and applied to the data assimilation system. The correction term to $\mathbf{x}_{j-1}^*$ measures the sensitivity of the operators definining the successive minimizations of the incremental VDA system to previous state estimate and indirectly to the observations.

Equation (5) can be used to provide an estimate of $\boldsymbol{\delta}\mathbf{K}_j$, using finite differences to approximate the inner-most term in the sum. However, like for the approximation of the product of a Hessian by a vector, using finite differences instead of a second order adjoint, the accuracy of the approximation depends on the amplitude of the finite difference and cannot be controlled. Moreover, this approximation cannot be used as a starting point to derive an approximate adjoint because some dependences are hidden. Thus, it is not possible to estimate the adjoint without developing the second -order adjoint, and its adjoint! This would be a daunting task for any meaningful data assimilation system, both in terms of development and computational costs.

The relative amplitudes of the correction terms with respect to the expressions given in Sections 3.1 and 3.2 depend on the degree of non-linearity of the operators **B** and **H** on the state estimate. For example, the observation operators for some observations comprise only interpolations and do not depend on the state estimate. In that case, the derivatives are zero, and there is no additional term. In other cases, the state estimate is used but the sensitivity are not known. In the case of 4D-Var, particularly strong constraint 4D-Var, the model is embedded in the observation operator $\mathcal{H}$, and we can expect more non-linear impact than in 3D-Var or in weak constraint 4D-Var. And, as described in Section 2, other constraints, such as balance constraints, can also be built into the **B** matrix. These constraints might be state dependent and would make **B** derivatives non-zero, although their sensitivity remains unknown.

If the overall sensitivity to the linearization state is weak, outer loop iterations would probably be unnecessary. Conversely, if the state dependence is strong, outer loops are necessary and the information brought by the second-order adjoint also becomes important. All current operational VDA systems use several outer iterations and have demonstrated benefits from these iterations, indicating that accounting for the second-order adjoint information is likely to be important. As explained above, this is a complex and tedious task for a realistic system and will not be attempted here. This caveat should be kept in mind when interpreting observation sensitivity results in the incremental VDA context.

3.4. Observation impact

3.4.1. First-order observation impact. Evaluation of sensitivity to observations of the analysis or of quantities related to forecasts issued from the analysis, requires the computation of $\partial\mathbf{x}_a/\partial\mathbf{y}$. Another application related to observation sensitivity studies is the study of observation impact. In that case, the goal is to estimate the impact observations had on a given measure of a forecast or analysis aspect. One measure of the impact of an observation on a given forecast aspect F can be defined as the difference between the values of F computed with and without taking that observation into account (This is the impact as measured by OSEs). Alternatively, assuming that $\delta\mathbf{x}$ is the increment generated by some observations, a measure of impact for these observations can also be defined as the variation of F due to that increment. An approximation of this measure of impact can be computed as a first-order Taylor series:

$$I_1 = \delta F = \langle \mathbf{g}, \delta\mathbf{x} \rangle,$$

where **g** is the gradient of F at the background point. Although this first-order approximation is in general not appropriate, it is examined first to demonstrate the influence of the VDA outer loops on observation impact.

For a linear data assimilation system or for the first outer loop of a non-linear incremental system, the increment can be written as a linear function of the departures (or innovations) from observations. As a result, the increment can be written as the sum of partial increments generated by each observation, and the impact of observations can be evaluated through the partial increment they generated. More precisely, the impact of the observations can be written as

$$I_1 = \langle \mathbf{g}, \delta\mathbf{x} \rangle = \langle \mathbf{g}, \mathbf{K}\mathbf{d} \rangle = \langle \mathbf{K}^T\mathbf{g}, \mathbf{d} \rangle.$$

The fact that the impact can be expressed as the scalar product of a constant quantity by the departure vector **d** has made observation studies popular since, once the vector on the left-hand side of the scalar product has been computed, the impact of any subset of observations can be computed by summing only over those components of **d**. As shown by the expression above, in a linear data assimilation system, the vector on the left-hand side is the observation sensitivity.

In a non-linear context, there is no agreed method to evaluate observation impact. Generalizing the method used in the linear case, it is possible to obtain an approximation of the increment as

$$\delta\mathbf{x} \approx \left[\frac{\partial G}{\partial \mathbf{y}}\right]_{\mathbf{x}_b} \mathbf{d} = \mathbf{K}_1\mathbf{d},$$

which leads to

$$I_1 \approx \langle \mathbf{K}_1^T\mathbf{g}, \mathbf{d} \rangle.$$

This is equivalent to measuring the observation impact in the first outer loop. Note that, contrary to the observation sensitivity approximation, the observation impact approximation is based on the adjoint of the first outer loop iteration. However, accounting for non-linearities through the outer loop iteration in 3D-Var or 4D-Var is beneficial to the quality of the analysis and ensuing forecast. Thus, it is important to quantify the impact of observations throughout the whole data assimilation process.

In an incremental VDA system, the impact can be defined as the sum of the impacts in each linear inner problem, using the fact that the total increment is the sum of the increments resulting from each minimization. However, as showed by eq. (3), the partial increments are a linear function of the departure vectors, only in the first outer loop iteration. From eq. (3), the total increment after iteration j is

$$\mathbf{x}_j - \mathbf{x}_b = \mathbf{K}_j\mathbf{d}_j + \mathbf{K}_j\mathbf{H}_j(\mathbf{x}_{j-1} - \mathbf{x}_b).$$

Applying this result recursively leads to

$$\mathbf{x}_a - \mathbf{x}_b = \textstyle\sum_{j=1}^{m} \mathbf{L}_j\mathbf{K}_j\mathbf{d}_j \quad \text{with}$$
$$\mathbf{L}_j = \mathbf{K}_m\mathbf{H}_m \ldots \mathbf{K}_{j+1}\mathbf{H}_{j+1} \quad \text{and} \quad \mathbf{L}_m = \mathbf{I}.$$

The total increment can be written as a linear combination of the observation departures from the various intermediate state estimates, and the impact of observations can be expressed as the scalar product:

$$I_1 = \langle \mathbf{g}, \mathbf{x}_a - \mathbf{x}_b \rangle = \sum_{j=1}^{m} \langle \mathbf{K}_j^T\mathbf{L}_j^T\mathbf{g}, \mathbf{d}_j \rangle.$$

The vectors that multiplies $\mathbf{d}_j$ in the expression above can be obtained in reverse order, using the fact that $\mathbf{L}_{j-1}^T = \mathbf{H}_j^T\mathbf{K}_j^T\mathbf{L}_j^T$. Thus, each adjoint operator is needed only once to compute the observation impact and the adjoints are applied in reverse order.

The computation of observation impact bears some similarities with the computation of observation sensitivities. As explained above, if the data assimilation system is linear, there is a simple relation between the two. If the system is non-linear, they both involve the same adjoint operators, but they are not as directly related. Observation sensitivity is a derivative, which can be computed using the adjoint technique. In the observation impact computation, the analysis increment is written as the sum of partial increments generated by each observation (Assuming such independent increments exist). Once this has been achieved, the adjoint operator is used to express the impact as a scalar product of, on one hand, the departure vectors and, on the other hand, a vector that will be considered independent of the observation values for the purpose of computing partial sums. This particular form of the impact is what allows computing the impact of any observation combination at the cost of a scalar product.

3.4.2. Second-order observation impact. Depending on the form of F, and in particular in the case where F is quadratic, the first-order estimate examined above might not be accurate. Even with a linear data assimilation system, second- or third-order Taylor series approximations are required, as shown by Errico (2007), Gelaro et al. (2007) and Trémolet (2007a). A second-order estimate of observation impact is thus given by the Taylor series:

$$I_2 = \mathbf{g}^T\delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T\mathbf{C}\delta\mathbf{x},$$

where **C** is the Hessian of the measure F. The first term of this approximation has been examined above. To be correct, the estimate of observation impact should be written as the second-order Taylor series with respect to the observation departures instead of $\delta\mathbf{x}$. This involves the second-order derivatives of G, which, as described in Section 3.3, cannot be computed in practice. However, contrary to F, which is not well approximated by a first-order approximation, it is expected that the first-order approximation of G is meaningful. Thus, the term involving second-order derivatives of G will be omitted in the following and $\delta\mathbf{x}$ is expressed as a linear combination of the successive observation departures, leading to a quadratic expression. To make observation impact studies practical, the quadratic term in the expression above must be written as

$$\delta\mathbf{x}^T\mathbf{C}\delta\mathbf{x} = \sum_{j=1}^{m} \langle \mathbf{K}_j^T\mathbf{L}_j^T\mathbf{C}\delta\mathbf{x}, \mathbf{d}_j \rangle.$$

This term can be combined with the first-order term so that each adjoint operator only needs to be applied once to compute one left-hand side vector for the scalar product. As explained earlier, the impact of a set of observations is obtained by summing the scalar product only over the components of interest of the departure vectors. However, in this quadratic term, the contributions of the other observations are still present in the other argument of the scalar product. This amounts to using the product $\delta\mathbf{x}^T\mathbf{C}\delta\mathbf{x}_s$, where $\delta\mathbf{x}$ is the full increment and $\delta\mathbf{x}_s$ is the increment generated by the selected observations when, in fact, $\delta\mathbf{x}_s^T\mathbf{C}\delta\mathbf{x}_s$ should be used.

If the observations of interest generate a partial increment that is orthogonal to the increment generated by the other observations (for the scalar product defined by **C**), the result will not be affected by this approximation. In practice, the matrix **C** is often diagonal, which means that increments will be orthogonal if they have components in different locations or for different variables. But, if two observations generate increments along the same directions, the result might be affected. Several scenarios can occur

(1) The two observations support each other and generate a similar increment: the result is unchanged.

(2) The observation of interest dominates the increment (because it is more accurate and given more weight in the analysis): the result is unchanged.

(3) The observation of interest has less impact on the increment (because it is less accurate and given less weight in the analysis): the diagnosed impact will reflect the larger increment generated by the other observation and be overestimated.

(4) The two observations have similar weight but generate increments of opposite signs which cancel each other: the diagnosed impact will reflect the small combined increment and be underestimated.

In practice, Gelaro et al. (2007) have observed that, for subsets containing large numbers of observations, the global results do not appear to be affected by this type of problems.

There are other caveats associated with this method for computing observation impact. Especially, in a non-linear data assimilation system, there are many complex interactions between the influence of various observations. For example, all observations have an impact on the successive state estimates that define the intermediate trajectories and thus the operators used in the incremental VDA system. When forecast aspects are measured, the forecast model introduces other higher-order terms in the observation impact approximation as shown by Errico (2007) and Gelaro et al. (2007). Finally, when a second- or higher-order estimate is used, the impact can still be written as the scalar product of the increment $\delta\mathbf{x}$ times a vector, but in this case, this vector cannot be interpreted as a derivative, it is not the observation sensitivity, and it is not independent of $\delta\mathbf{x}$. It should only be interpreted as a vector of weights given to each observation.

An accurate computation of observation impact should account for all these higher order effects, but as shown above in the sections devoted to observation sensitivity computations, these are very difficult to estimate. It is expected that the results obtained by computing the left-hand side weights for the scalar product once and summing only over the relevant components of the departure vectors, are nevertheless useful in practice, in particular, when averaged over large enough subsets of observations, even though all the pitfalls mentioned above should be considered when interpreting observation impact results.

4. Computing the adjoint of the inner loop

4.1. General observation sensitivity algorithm

In this section, we assume that the data assimilation system is linear. This hypothesis has been made in most observation sensitivity studies. It also corresponds to an incremental variational system with one outer loop iteration and can be used as a building block to construct the adjoint of a complete incremental system with several outer loop iterations, as described in Section 3. For clarity of the presentation, we omit the outer loop index j throughout this section.

As shown in Section 3, the remaining difficulty in obtaining the sensitivity to observations is the computation of $\mathbf{K}^T$. For convenience, we introduce the matrix:

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \text{ i.e. } \mathbf{K} = \mathbf{A}^{-1}\mathbf{H}^T\mathbf{R}^{-1}.$$

A is the Hessian of the variational cost function J, and if the operators are linear and the data assimilation system is optimal, **A** is also the inverse of the analysis error covariance matrix.

From the definition of **A**, we get $\mathbf{K}^T = \mathbf{R}^{-T}\mathbf{H}\mathbf{A}^{-T}$. However, since **B** and **R** are covariance matrices, they are self adjoint. As a consequence, **A**, and its inverse, are self adjoint and $\mathbf{K}^T = \mathbf{R}^{-1}\mathbf{H}\mathbf{A}^{-1}$.

The operators $\mathbf{R}^{-1}$ and **H** are already available in the VDA system, thus, the only difficulty is to compute the product of a vector by $\mathbf{A}^{-1}$ or to solve the linear system generated by the matrix **A**.

Taking the derivative of eq. (2), the gradient of the quadratic cost function being minimized in the inner loop of the VDA system is

$$\nabla J = \mathbf{B}^{-1}(\delta\mathbf{x} - \mathbf{b}) + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}).$$

Minimizing the quadratic cost function J is equivalent to solving the equation $\nabla J = 0$ which, with the notations introduced above, is the linear system

$$\mathbf{A}\delta\mathbf{x} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} + \mathbf{B}^{-1}\mathbf{b}. \tag{6}$$

This shows that the minimization performed in the inner loop is equivalent to solving a linear system generated by the matrix **A**. Thus, the linear system that has to be solved to compute $\mathbf{K}^T$, is already being solved, with a different right-hand side, in the data assimilation process. The basis for computing $\mathbf{K}^T$ is to use the same algorithm, replacing the right-hand side where necessary. Two possibilities for inverting **A** are described in the next two paragraphs. After this linear system is solved, the solution is multiplied by $\mathbf{R}^{-1}\mathbf{H}$ to obtain the product by $\mathbf{K}^T$ and the contribution to the sensitivity of $\mathcal{F}$ with respect to observations or to the observation impact.

4.2. Inversion by minimization

Many algorithms can be used to solve the linear system resulting from the incremental VDA inner loop minimization. As the matrices involved in operational data assimilation cannot be stored, operators giving the value of the cost function and its gradient at a given point are usually the only tools available. Given this limitation, conjugate gradient and quasi-Newton algorithms are the most commonly used to solve the VDA problem.

For example, the conjugate gradients algorithm to solve the linear system $\mathbf{Ax} + \mathbf{g} = 0$ requires the ability to compute the product of the matrix $\mathbf{A}$ by the vector $\mathbf{x}$. In practice, the only operator easily available gives the gradient of the VDA cost function, which, in terms of linear system, is also $\mathbf{Ax} + \mathbf{g}_0$, where $\mathbf{g}_0$ is the gradient computed at the origin. The gradient can be computed with $\mathbf{d} = 0$ and $\mathbf{b} = 0$ in eq. (6), which gives the product $\mathbf{Ax}$ directly where needed in the minimization algorithm. Or, the gradient of the data assimilation cost function in the direction $\mathbf{x}$ can be computed without modification, and the original gradient at the origin $\mathbf{g}_0$ can be subtracted from it. Although there is no significant difference in the computational costs of these methods, ease of implementation might make one or the other more suitable in a given system.

In quasi-Newton algorithms, the problem is treated as a minimization problem rather than a linear system, and the full gradient of the function to be minimized is required. Adapting the minimization algorithm to compute the observation sensitivity comprises modifying the definition of the cost function to accommodate the new right-hand side. Again, this only requires subtracting the gradient of the original VDA cost function at the origin and adding the new right-hand side at the end of each gradient evaluation. This requires only very limited modification of the data assimilation code.

This procedure implies that the linear system is solved to convergence so that the product by $\mathbf{A}^{-1}$ is effectively obtained. The cost of the minimization should be the same as the cost of the minimization of the direct VDA cost function, since both problems have the same conditioning, and more than that, their entire spectrum is the same. It is similar to algorithms such as the one developed by Baker and Daley (2000) in observation space data assimilation systems or Cardinali and Buizza (2004) and Cardinali (2008) in 4D-Var.

4.3. Adjoint of the approximate Hessian inverse

The method described above is valid at convergence of both the direct and transpose minimizations. However, if a conjugate gradient or a related minimization algorithm is used, at any given iteration i of the inner loop, the estimate of the solution by the minimization algorithm takes the form

$$\delta \mathbf{x}_i = \widetilde{\mathbf{A}}_i \mathbf{g}_0,$$

where $\widetilde{\mathbf{A}}_i$ represents the combined effect of the accumulated iterations up to that point. As the number of iterations increases, the matrix $\widetilde{\mathbf{A}}_i$ becomes a better approximation of $\mathbf{A}^{-1}$, and the iterates converge towards the analysis increment.

Moreover, when the Lanczos algorithm is used, the matrix $\widetilde{\mathbf{A}}_i$ can be formed explicitly and takes the particular form

$$\widetilde{\mathbf{A}}_i = \mathbf{Q}_i (\mathbf{L}_i \mathbf{D}_i \mathbf{L}_i^T)^{-1} \mathbf{Q}_i^T,$$

where $\mathbf{Q}_i$ is the matrix formed by the i Lanczos vectors and $\mathbf{D}_i$ and $\mathbf{L}_i$ are rank i matrices, respectively, diagonal and lower bi-diagonal. (See, for example, Golub and Van Loan, 1996 for details.) Because $\mathbf{D}_i$ and $\mathbf{L}_i$ are small matrices, and because of their particular sparse structure, the product $\mathbf{L}_i \mathbf{D}_i \mathbf{L}_i^T$ can very easily be inverted and at very low cost. Thus, $\widetilde{\mathbf{A}}_i$ can easily be accessed and applied to any vector.

This specificity of the Lanczos algorithm opens the way for another method for computing the sensitivities to observations. In practice, after i iterations of the minimization algorithm, we have not computed the increment implied by eq. (3):

$$\delta \mathbf{x} = \mathbf{A}^{-1}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} + \mathbf{B}^{-1} \mathbf{b}),$$

but only an approximation of it

$$\delta \mathbf{x}_i = \widetilde{\mathbf{A}}_i (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} + \mathbf{B}^{-1} \mathbf{b}).$$

We define

$$\widetilde{\mathbf{K}}_i = \widetilde{\mathbf{A}}_i \mathbf{H}^T \mathbf{R}^{-1}.$$

This partial Kalman gain matrix can easily be transposed using the decomposition of $\widetilde{\mathbf{A}}_i$ described above if the Lanczos algorithm is used in the minimization. Thus, one can obtain the sensitivity to observations of the partial increment at iteration i of the inner loop minimization. At the end of the minimization, after n iterations, the solution is multiplied by $\mathbf{R}^{-1}\mathbf{H}$ to obtain the sensitivity of $\mathcal{F}$ with respect to the observations:

$$\frac{\partial \mathcal{F}}{\partial \mathbf{y}} = \widetilde{\mathbf{K}}_n^{\,T} \frac{\partial \mathcal{F}}{\partial \mathbf{x}_\mathrm{a}}.$$

If the minimization has properly converged, this should yield the same result as the method described in Section 4.2. However, regardless of the convergence or non-convergence of the solution of the linear system, this algorithm always generates the adjoint of the gain matrix actually used in the assimilation. If the sensitivity to observations is to reflect the operator effectively used in the data assimilation process, this algorithm allows the use of an exact adjoint of this operator. Other potential uses for this algorithm are described in Section 5.3.

5. Numerical results

5.1. Adjoint validation method

For faster convergence of the data assimilation algorithm, the minimization of the inner loop cost function, defined by eq. (2), is usually pre-conditioned. To that effect, a change of variable is applied and the cost function is re-written as a function of the new control variable χ defined by $\delta\mathbf{x} = \mathbf{E}\chi$, where $\mathbf{E}$ is the pre-conditioner:

$$J(\chi) = \frac{1}{2}(\mathbf{E}\chi - \mathbf{b}_j)^T\mathbf{B}^{-1}(\mathbf{E}\chi - \mathbf{b}_j) + \frac{1}{2}(\mathbf{H}_j\mathbf{E}\chi - \mathbf{d}_j)^T \mathbf{R}^{-1}(\mathbf{H}_j\mathbf{E}\chi - \mathbf{d}_j).$$

The pre-conditioner $\mathbf{E}$ is chosen as a square root of the background error covariance matrix, such that $\mathbf{E}\mathbf{E}^T = \mathbf{B}$. With that choice, the gradient of the cost function in control space is

$$\nabla J(\chi) = \chi - \mathbf{E}^{-1}\mathbf{b}_j + \mathbf{E}^T\mathbf{H}_j^T\mathbf{R}^{-1}(\mathbf{H}_j\mathbf{E}\chi - \mathbf{d}_j).$$

The solution of the pre-conditioned problem is the solution of the linear system

$$(\mathbf{I} + \mathbf{E}^T\mathbf{H}_j^T\mathbf{R}^{-1}\mathbf{H}_j\mathbf{E})\chi = \mathbf{E}^T\mathbf{H}_j^T\mathbf{R}^{-1}\mathbf{d}_j + \mathbf{E}^{-1}\mathbf{b}_j.$$

The form of this problem implies that all its eigenvalues are greater than or equal to 1. Furthermore, since the size of the control variable (and of the identity matrix in the system above) is at least one order of magnitude larger than the number of observations, it means that the matrix to be inverted has a majority of eigenvalues equal to 1, which makes this pre-conditioner very efficient. $\mathbf{E}$ might not be invertible, which would make the term $\mathbf{E}^{-1}\mathbf{b}_j$ problematic. However, this term can usually be computed, without explicitly applying $\mathbf{E}^{-1}$, as a by-product of the minimization algorithm—it is the estimate, in control space, of the solution at the end of the previous minimization: $\mathbf{E}^{-1}\mathbf{b}_j = -\mathbf{w}_{j-1}$, with $\mathbf{w}_j = \sum_{k=1}^{j}\chi_k$.

We define $\widehat{\mathbf{A}_j}$ as the inverse of the Hessian for the pre-conditioned system

$$\widehat{\mathbf{A}_j} = (\mathbf{I} + \mathbf{E}^T\mathbf{H}_j^T\mathbf{R}^{-1}\mathbf{H}_j\mathbf{E})^{-1}.$$

The solution in the pre-conditioned space is

$$\chi_j = \widehat{\mathbf{K}_j}\mathbf{d}_j - \widehat{\mathbf{A}_j}\mathbf{w}_{j-1}, \quad (7)$$

where $\widehat{\mathbf{K}_j} = \widehat{\mathbf{A}_j}\mathbf{E}^T\mathbf{H}_j^T\mathbf{R}^{-1}$. Taking the scalar product of the last equation by $\partial\mathcal{F}_j/\partial\chi_j$ gives

$$\left\langle\frac{\partial\mathcal{F}_j}{\partial\chi_j}, \chi_j\right\rangle = \left\langle\frac{\partial\mathcal{F}_j}{\partial\chi_j}, \widehat{\mathbf{K}_j}\mathbf{d}_j\right\rangle - \left\langle\frac{\partial\mathcal{F}_j}{\partial\chi_j}, \widehat{\mathbf{A}_j}\mathbf{w}_{j-1}\right\rangle.$$

We now define the measure $\widehat{\mathcal{F}}$ as half the squared norm of the increment generated by outer loop j in the norm associated to the background error covariance matrix

$$\widehat{\mathcal{F}_j}(\delta\mathbf{x}_j) = \frac{1}{2}\langle\delta\mathbf{x}_j, \mathbf{B}^{-1}\delta\mathbf{x}_j\rangle = \frac{1}{2}\langle\chi_j, \chi_j\rangle.$$

Using the definition of an adjoint and the fact that $\widehat{\mathbf{A}_j}$ is self-adjoint leads to

$$\langle\chi_j, \chi_j\rangle = \left\langle\widehat{\mathbf{K}_j}^T\frac{\partial\widehat{\mathcal{F}_j}}{\partial\chi_j}, \mathbf{d}_j\right\rangle - \langle\widehat{\mathbf{A}_j}\chi_j, \mathbf{w}_{j-1}\rangle.$$

Thus,

$$\langle\chi_j, \chi_j\rangle = \left\langle\frac{\partial\widehat{\mathcal{F}_j}}{\partial\mathbf{y}}, \mathbf{d}_j\right\rangle - \langle\widehat{\mathbf{A}_j}\chi_j, \mathbf{w}_{j-1}\rangle. \quad (8)$$

These scalar products can be computed and compared. The term $\widehat{\mathbf{A}_j}\chi_j$ is an intermediate result, available during the computation of $\partial\widehat{\mathcal{F}_j}/\partial\mathbf{y}$ and does not introduce any additional difficulty or cost to the computation. In a linear VDA system, or in the first minimization of a non-linear VDA system, $\mathbf{b}_j = 0$ and the formula simplifies to

$$\langle\chi_j, \chi_j\rangle = \left\langle\frac{\partial\widehat{\mathcal{F}_j}}{\partial\mathbf{y}}, \mathbf{d}_j\right\rangle.$$

This is the usual adjoint test applied to the adjoint of the data assimilation system. This formula has been used by Zhu and Gelaro (2008) to validate sensitivity calculations based on the line by line adjoint of the GSI data assimilation system.

5.2. *Implementation and numerical adjoint validation*

The data assimilation system used for the experiments presented here is the grid-point statistical interpolation (GSI) system (Wu et al. (2002)). This is a 3D-Var system, currently used operationally at NCEP and GMAO. A 4D-Var version of the GSI is currently being developed at GMAO. The implementation of the algorithms described in this paper is straightforward and independent of the definition of the observation operator $\mathcal{H}$ and thus of the three or four-dimensional aspects of the VDA. More accurately, since the adjoint of the VDA system makes use of the operators used in the direct VDA system, the observation sensitivity algorithm inherits all the changes made in the direct VDA system. This aspect was one of the primary motivations for this work, since writing the adjoint code of the 4D-Var version of the GSI would require major developments, even though the adjoint of the 3D-Var system had already been developed (Zhu and Gelaro, 2008).

The validation tests described in the previous section have been performed in both the 3D-Var and 4D-Var versions of the GSI. Examples of numerical results obtained, when the adjoint is computed by replacing the right-hand side term in the minimization (Section 4.2), are given in Table 1 where, based on equation (8), relative error is defined as

$$e = \left|\langle\partial\widehat{\mathcal{F}_j}/\partial\mathbf{y}, \mathbf{d}_j\rangle - \langle\widehat{\mathbf{A}_j}\chi_j, \mathbf{w}_{j-1}\rangle - \langle\chi_j, \chi_j\rangle\right| / \langle\chi_j, \chi_j\rangle.$$

The table shows that the results depend on the convergence of the minimizations and that the accuracy improves with the number inner loop iterations. The results are in agreement with results obtained by Xu et al. (2006), who implemented a similar algorithm in an observation space VDA system and showed that the same test was accurate to five digits. This is better than the accuracy obtained by Zhu and Gelaro (2008) in their validation of the adjoint code of the 3D-Var GSI (errors of the order of a few percents, Yanqiu Zhu, personal communication, 2007) because some parts of the code were ignored or simplified in the development of the tangent linear system upon which the line by line adjoint is based. In the present case, because the code of the direct assimilation system is used, there are no such discrepancies.

Table 1. Validation test of the adjoint of the 3D-Var and 4D-Var GSI obtained by minimization (algorithm described in Section 4.2). The number of inner iterations n is the same for the direct and adjoint minimizations in each case

	n	j	$\langle \partial \mathcal{F}_j / \partial \mathbf{y}, \mathbf{d}_j \rangle$	$\langle \widehat{\mathbf{A}_j} \chi_j, \chi_{j-1} \rangle$	$\langle \chi_j, \chi_j \rangle$	Relative error
		1	3.802×10^4	0.0	3.507×10^4	8.4×10^{-2}
3D-Var	50	2	2.742×10^4	4.535×10^3	2.215×10^4	3.3×10^{-2}
		3	1.951×10^4	4.699×10^3	1.444×10^4	2.6×10^{-2}
		1	4.493×10^4	0.0	4.482×10^4	2.3×10^{-3}
3D-Var	100	2	2.473×10^4	3.234×10^3	2.149×10^4	3.5×10^{-4}
		3	1.825×10^4	3.460×10^3	1.479×10^4	2.9×10^{-4}
		1	4.150×10^4	0.0	3.853×10^4	7.7×10^{-2}
4D-Var	50	2	2.767×10^4	1.781×10^3	2.538×10^4	2.0×10^{-2}
		3	1.918×10^4	1.186×10^3	1.764×10^4	2.0×10^{-2}
		1	4.775×10^4	0.0	4.770×10^4	1.0×10^{-3}
4D-Var	100	2	2.437×10^4	-1.971×10^3	2.634×10^4	8.3×10^{-5}
		3	1.847×10^4	-1.193×10^2	1.859×10^4	7.4×10^{-5}

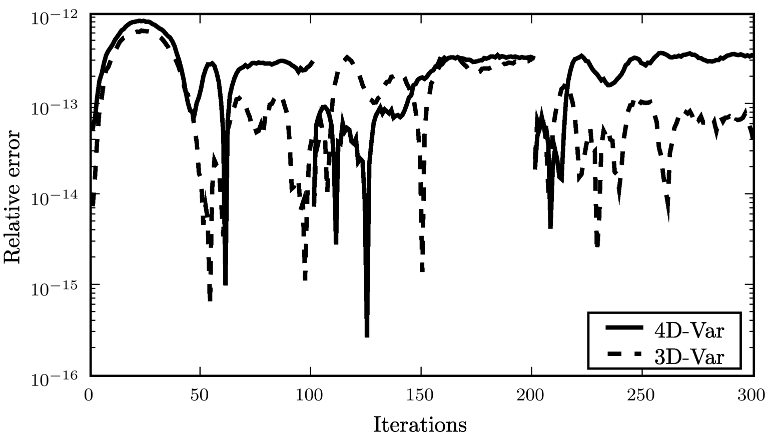

Fig 1. Validation test of the adjoint of the 3D-Var and 4D-Var GSI obtained using the adjoint of the approximate inverse of the Hessian (algorithm described in Section 4.3). The results are shown for three outer loop iterations with 100 inner iterations in each minimization.

A property of the algorithm comprising the transposition of the approximate gain matrix (Section 4.3) is that it can be applied at any stage during the minimization process (This is also true of a line by line adjoint). In this case, the validation test described above is valid at every iteration during the inner loop minimization. The relative error achieved in 3D-Var and 4D-Var at every iteration is shown in Fig. 1. In this implementation, the accuracy is of the order of 12 to 13 digits. This result conforms to expectations, since in this case, the adjoint is the exact adjoint (to machine precision) of the linear operator effectively used within each iteration and is independent of the convergence of the minimization.

5.3. Observation impact during the minimization

The observation space scalar product $\langle \partial \widehat{\mathcal{F}_j} / \partial \mathbf{y}, \mathbf{d}_j \rangle$ used in the validation of the adjoint can also be interpreted as the contribution of observations to the current partial increment. It is related to the observation impact on the measure J_b, considering the impact only in the current minimization. Like other measures of observation impact, it can be computed for any subset of observations. And, when the adjoint of the data assimilation system is computed as described in Section 4.3, this contribution can be computed at every iteration of the inner loop.

This provides a measure of the observation usage as the minimization progresses and provides useful diagnostics of the data assimilation process. Figure 2 shows this observation partial impact as the minimization progresses in 3D-Var and 4D-Var. In addition to the diagnosis of the observation usage, this type of plot can provide useful information about the data assimilation system. For example, during the development of the 4D-Var version of the GSI, such plots showed several abnormal behaviours, such as the non-convergence of one of the curves, which pointed to errors in the handling of a particular observation type that would have taken much longer to identify otherwise.

It can be seen from the figure that the impact of radiance data is generally larger in 4D-Var for all iterations. This is expected since radiance data is evenly distributed in time, and 4D-Var should be able to make better use of asynoptic data than 3D-Var. The figure also shows that the curves tend to flatten sligthly earlier in the second and third minimizations in 3D-Var than in 4D-Var. This might be related to the results of the next section,

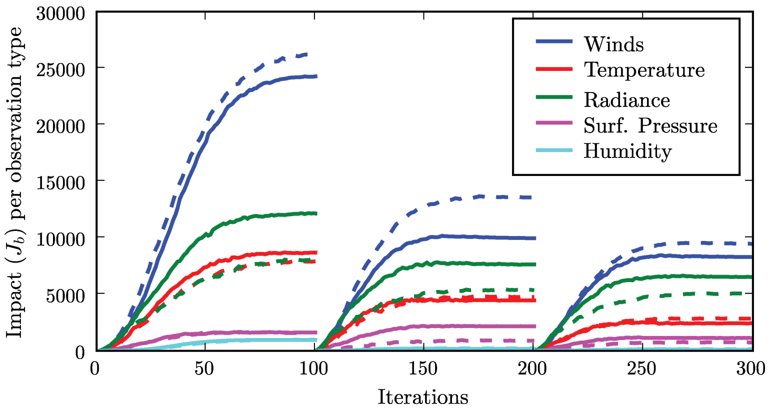

Fig 2. Partial impact of observations in inner loop iterations in 4D-Var (plain lines) and 3D-Var (dashed lines) with a 6-h assimilation window. The results are shown for three outer loop iterations with 100 inner iterations in each minimization.

which indicate that outer loop iterations are less important in 3D-Var than in 4D-Var and implies that it would require less iterations in the inner loop to reach convergence and could be used to define a stopping criterion for the minimization. This measure of observation impact is related to the notion of degrees of freedom for signal (DFS) as presented by Fisher (2003).

This technique does not give directly the total observation impact unless the data assimilation system comprises only one outer iteration. Adding the impacts from all the outer loops obtained in this manner is different from the total impact for the reasons discussed in Section 3.4.

5.4. *Observation impact with several outer loops*

Following the results given in Section 3.4, the second-order approximation of observation impact on a measure F can be written as the sum of scalar products of weight vectors with the observation departures from the successive outer loop state estimates. As in the previous paragraphs, and for illustrative purposes, J_b is used as the measure for which observation impact is computed. This choice is motivated by the fact that J_b can be interpreted as a measure of the information that has been transferred from the observations to the analysis. For this choice of analysis aspect measure, the first order approximation of the sensitivity to the analysis $\partial F/\partial \mathbf{x}_a$ is zero. This illustrate the necessity to use at least second-order approximations for observation impact computations, as shown by Gelaro et al. (2007) and Trémolet (2007a), even for measures of aspects other than forecast error. Accounting for the pre-conditioning, the increment is given by eq. (7). Applying the result given by this equation recursively gives

$$\mathbf{w}_m = \sum_{j=1}^{m} \widehat{\mathbf{L}_j}\widehat{\mathbf{K}_j}\mathbf{d}_j,$$

where $\widehat{\mathbf{L}_j} = (\mathbf{I} - \widehat{\mathbf{A}_m}) \cdots (\mathbf{I} - \widehat{\mathbf{A}_{j+1}})$ and $\widehat{\mathbf{L}_m} = \mathbf{I}$.

Given the particular form of J_b, the second-order approximation of observation impact is $I_2 = 1/2 < \mathbf{w}_\mathrm{a}, \mathbf{w}_\mathrm{a} >$ where $\mathbf{w}_\mathrm{a} = \mathbf{w}_m$ is the total increment in control space; thus,

$$I_2 = \frac{1}{2}\sum_{j=1}^{m} \langle \mathbf{w}_\mathrm{a}, \widehat{\mathbf{L}_j}\widehat{\mathbf{K}_j}\mathbf{d}_j \rangle,$$

which leads to the second-order approximation of the observation impact on J_b:

$$I_2 = \frac{1}{2}\sum_{j=1}^{m} \langle \mathbf{R}^{-1}\mathbf{H}_j\mathbf{E}\widehat{\mathbf{A}_j}\mathbf{z}_j, \mathbf{d}_j \rangle,$$

with

$$\mathbf{z}_m = \mathbf{w}_\mathrm{a} \text{ and } \mathbf{z}_{j-1} = (\mathbf{I} - \widehat{\mathbf{A}_j})\mathbf{z}_j.$$

The product $\widehat{\mathbf{A}_j}\mathbf{z}_j$ can be computed once and used to compute the contribution to the impact and the input to the adjoint of the previous minimization. This formulation is more efficient than the formulation involving $\mathbf{H}_j^T\mathbf{K}_j^T$ given in Section 3.4 and can be used for any measure which gradient can be expressed in control space.

The observation impact of several observation types in 3D-Var and 4D-Var with 1, 2 and 3 outer loop iterations are shown in Fig. 3. Two main results can be observed on the figure. First, the outer loop iterations seems to have more impact in 4D-Var than in 3D-Var. This corresponds to the fact that non-linearities from the outer loop affect the 4D-Var minimization through the trajectories for the tangent linear and adjoint models much more than in 3D-Var, where the only non-linearities come from the instantaneous observation operators and do not evolve throughout the assimilation window. The observation operators that rely the most on the state estimate are for radiance data through the radiative transfer model, in agreement with the figure which shows an impact of the outer loop for radiance data even in 3D-Var.

The second result highlighted by Fig. 3 is that the overall observation impact is smaller in 4D-Var than in 3D-var. This is obtained despite the fact that more observations where used in 4D-Var, as shown by Table 2. The table also shows that the initial value of J_o was lower in 4D-Var, despite the fact that more observations were used. This shows that the analyses produced by

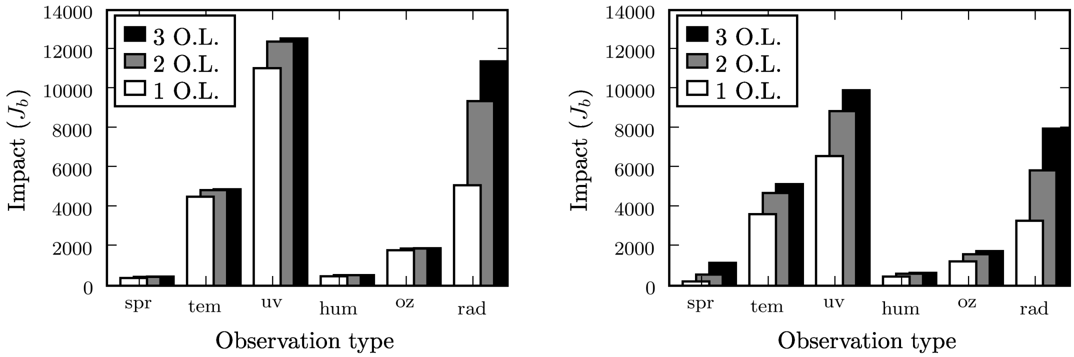

Fig 3. Observation impact per observation type for 3D-Var (left-hand side) and 4D-Var (right-hand side) with 1, 2 and 3 outer loop iterations.

Table 2. Initial J_o values for the 3D-Var and 4D-Var experiments corresponding to the impact presented on Fig. 3

	n_{obs}	J_o	J_o/n_{obs}
3D-Var	813 217	807 297.1	0.993
4D-Var	968 216	679 954.3	0.702

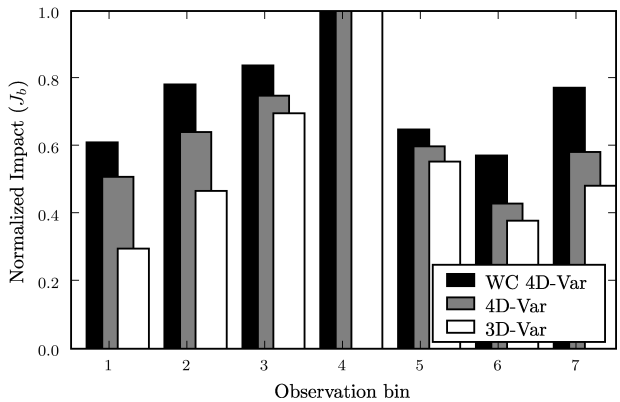

Fig 4. Average impact per observation for each bin within the 6-h assimilation window with 3D-Var, 4D-Var and weak constraint 4D-Var. The impacts are normalized so that the impact in the bin centred around the synoptic time is 1.

the 4D-Var experiment are better, providing a better background for the following cycle, thus requiring smaller corrections in subsequent analyses.

This points to another caveat of observation impact studies: observations are often considered valuable if they have a large impact. The fact that observations have a small impact could indicate that the data assimilation system is not extracting enough information from the observations, but also that the background is good and requires little correction. However, even that does not mean that the observing system is of little value since other observations, usually from the same observing system, were used to estimate the background. Ideally, a very good data assimilation system would produce very accurate analysis, and thus, background states that require very small corrections. The impact of observations in a given data assimilation cycle might be very small but it would be incorrect to conclude that observations have little value since it is the impact of observations over many data assimilation cycles that determines the level of performance of the data assimilation system. To make a judgement about the overall value of the observing system would require running the adjoint of the whole cycling data assimilation system over long periods and considered as a single operator for which the input is the entire set of observations for that period and the output, the time-series of analyses. That should be done for a period of at least 10 d since this can be considered to be the period after which an atmospheric data assimilation system has forgotten the initial background, and thus, the observations prior to that, as shown by Fisher et al. (2005).

Although observation impact studies are useful to compare the impact of observations within a given data assimilation system, the result above shows that comparisons of impact in different data assimilation systems can be misleading, or at least difficult to interpret. Finally, the measure used here does not allow to conclude whether the impact of observations is beneficial or detrimental to the quality of the analysis. Forecast error impacts studies will be necessary to answer that question.

5.5. *Time within the assimilation window*

The observations can be gathered in groups other than observation type. For example, they can be grouped by time slots within the assimilation window. Figure 4 shows histograms of observation impact for time slots centred around each hour in the assimilation window for 3D-Var, 4D-Var and weak constraint 4D-Var (similar to the algorithm described by Trémolet, 2007c, with the model error covariance matrix proportional to **B**) and where, for illustrative purpose, the background term J_b of the cost function is chosen as the measure for which the impact of observations are computed in 3D-Var and 4D-Var. In weak constraint 4D-Var, the measure for which observation impact is computed also includes the model error term. Because of this and because of the difference between 3D-Var and 4D-Var described in the previous section, the impacts have been normalized so that the impact in the bin centred around the synoptic time is 1.

Figure 4 shows that, as the observations are further away from the synoptic time, they have relatively more impact in 4D-Var than in 3D-Var and even more impact in weak constraint 4D-Var. As 3D-Var does not take into account time information, it cannot make the most accurate use of data that is distributed in time; thus, the fact that observations away from the analysis time have less impact, is expected. In 4D-Var, the tangent linear and adjoint models propagate the information brought by asynoptic observation to the analysis time, thus giving more impact to these observations. But, 4D-Var is affected by model error, which deteriorates the propagation of information. Weak constraint 4D-Var, accounting for model error, should, in principle, be able to use all observations with the same efficiency. In practice, this is limited by the various simplifications, which are necessary to make weak constraint 4D-Var practical. We also point out that 4D-Var and weak constraint 4D-Var are still in development. The diagnostics proposed here should be considered as part of the tuning and validation; thus, these results should be considered as illustrative of these systems and not as final.

5.6. *Practical use and computational cost*

The practical use of the algorithms proposed in Section 4 depends on their cost, both in terms of computations and storage. The evaluation of observation impact and forecast sensitivity comprises several steps: the data assimilation is performed; the forecast model is integrated; the aspect measure and its gradient are computed; the adjoint model is integrated backwards to analysis time and, finally, the adjoint of the data assimilation system is applied. Some information from the forward data assimilation run is needed in the adjoint and must be either stored or recomputed. For example, in both algorithms, the linearization trajectories and the observations characteristics are needed. All the required observation related information is already archived in most data assimilation systems and can be re-used without difficulty. The linearization trajectories are usually not stored and can represent very large volumes of data. They can be stored or re-computed by re-running the outer loop forecast if the intermediate outer loop increments are saved.

The computational cost of the minimization-based algorithm is of the same order as that of the forward data assimilation process. The minimizations have the same condition number, and even more, their full spectrum is the same as the same matrix is inverted. Thus, the convergence rate should be identical and the minimizations have the same cost. Since the non-linear (high-resolution) evaluation of the observation operator is not necessary in the adjoint, the overall cost might be lower than that for the full direct assimilation, unless the trajectories are recomputed in which case the overall costs are similar.

The computational cost of the adjoint-based method is very low: less than the equivalent of one inner iteration. The linear algebra is less than in a direct inner iteration and the application of $\mathbf{R}^{-1}\mathbf{H}$ is approximately half of the cost of the gradient evaluation, which is performed at each inner iteration of the direct assimilation and involves $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$. Thus, the overall cost of the adjoint is, at most, of the order of a few percents of the cost of the direct assimilation. This algorithm uses the Lanczos vectors, which were computed in the direct assimilation. They can either be stored or re-computed at the cost of re-running the analysis, in which case, the cost of the adjoint of the assimilation operator is the same as the cost of the analysis itself. In the specific case where this algorithm is used to diagnose the sensitivity of the increment to observations, as was the case for some numerical results presented in this paper (Section 5.3), the adjoint can be computed immediately, during the minimization, at the cost of approximately half an inner iteration, and without any additional storage requirements. This makes it affordable to use as a systematic diagnostic of the data assimilation system if necessary.

Observation impact studies are related to OSE. One difference is that OSEs easily allow the estimation of one observing system impact on several forecast aspects measures, whereas observation impact studies can easily provide the impact of many observing system combinations on one forecast aspect measure. The algorithms presented in this paper allow the combination of both. After the forward analysis and forecast have been performed, several aspects of the forecast can be evaluated and their gradients produced. Then, the forecast adjoint integration has to be performed for each gradient, which can be expensive. However, once the gradients with respect to the analysis are known, and especially if the adjoint of the approximate Kalman gain matrix is applied through the use of the Lanczos vectors, the cost of applying the adjoint of the assimilation system to multiple gradients is low. Thus, observation impact can be computed for multiple forecast aspects and multiple observing systems at a reasonable cost.

6. Conclusions

We have shown how the outer loops of an incremental variational model space data assimilation system affect observation sensitivity and observation impact computations and how this effect could be accounted for. To fully account for this effect, the second-order adjoint of the data assimilation system is required, which is not practical for any meaningful system. This is unfortunate, since the outer loops of 4D-Var bring significant improvement to the analysis quality, which implies that the observations have a meaningful impact throughout all the outer iterations. Experiments in a simplified system would be of interest and should be pursued in the future to quantify the errors introduced by ignoring those higher-order terms. With some approximations in the treatment of outer loops, observation impact and observation sensitivity studies are, nevertheless, possible.

In that context, we have developed and tested two algorithms to compute the adjoint of the inner loop minimizations. The first algorithm, based on inverting the matrix $\mathbf{A}_j$ by way of using the minimization algorithm with a different right-hand side (Section 4.2), requires convergence of both the direct and adjoint minimizations. In practice, the minimizations don't always converge and the adjoint of the data assimilation system is not always exact. Moreover, the data assimilation process itself is prone to convergence issues (see Andersson et al., 2000 or Trémolet, 2007b for examples). Investigation of observations impact in these cases can be useful, for example, to understand the lack of convergence.

By contrast, the second algorithm, based on transposing the linear operator, effectively applied inside the minimization algorithm (Section 4.3), always yields the exact adjoint of the inner loop minimization. This algorithm is specific to the Lanczos minimization algorithm but can be applied at any stage during the minimization and regardless of convergence. This allows for the study of the convergence of the minimization algorithm in observation space, which is a new application. It also allows for the study of the impact of observations on the forecast issued from any specific analysis, which might be useful to diagnose bad forecasts, even in the case of a failure of the minimization. This is different from diagnostics that evalute statistical properties of observing systems and often rely on the assumptions that the data assimilation system is linear and optimal and is also a new application. The algorithm proposed here can be used both to diagnose the properties of the observing system and for diagnosing and understanding the data assimilation system itself, about which, it provides more relevant and detailed information.

Both algorithms immediatly reflect changes in the data assimilation system since they make use of the same operators and do not require additional maintenance. Some results in 3D-Var, 4D-Var and weak constraint 4D-Var were presented in this paper. Since the 4D-Var system is still being developed and validated, the results presented in this paper should not be considered as final, but as part of the validation process for this new data assimilation system. This would not be possible with a line by line adjoint since any modification or new development in the data assimilation system would require changes in the adjoint, which would be extremely tedious and costly and is another advantage of the approach proposed in this paper.

Many approximations are necessary to make observation impact computations practical and have been highlighted in this paper. Some might make observation impact studies difficult to interpret, and all these approximations should be considered, since in most cases, their actual influence is unknown. Studies in simplified setups, where all the higher-order terms could be computed, would be useful, for example, to examine the importance of the second-order adjoint information with multiple outer loop iterations. Comparison of results with OSE results might also be useful in that respect, for more complex systems.

7. Acknowledgments

The author would like to thank Ron Gelaro for his encouragement and support in pursuing this work, as well as for his helpful comments on an earlier version of this manuscript. Work with the GSI would not have been possible without invaluable advice and help from Ricardo Todling. Ron Errico also provided useful comments on this manuscript.

References

Andersson, E., Fisher, M., Munro, R. and McNally, A. 2000. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Q. J. R. Meteorol. Soc.* **126**, 1455–1472.

Baker, N. and Daley, R. 2000. Observation and background adjoint sensitivity in the adaptative observation targeting problem. *Q. J. R. Meteorol. Soc.* **126**, 1431–1454.

Cardinali, C. 2008. Monitoring the observation impact in the short-range forecast. *Q. J. R. Meteorol. Soc.* Submitted.

Cardinali, C. and Buizza, R. 2004. Observation sensitivity to the analysis and the forecast: a case study during ATreC targeting campaign.In: *Proceedings of the First THORPEX International Science Symposium*. 6-10 December 2004, Montreal, Canada , WMO TD 1237 WWRP/THORPEX No. 6.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

Errico, R. 2007. Interpretations of an adjoint-derived observational impact measure. *Tellus* **59A**, 273–276.

Fisher, M. 2003. Estimation of Entropy Reduction and Degrees of Freedom for Signal for Large Variational Analysis Systems. Tech. Memo. 397, ECMWF.

Fisher, M., Leutbecher, M. and Kelly, G. 2005. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **131**, 3235–3246.

Gelaro, R., Zhu, Y. and Errico, R. 2007. Examination of various-order adjoint-based approximations of observation impact. *Meteorol. Z.* **16**, 685–692.

Golub, G. and Van Loan, C. 1996. *Matrix Computations* 3rd Edition. Johns Hopkins University Press, Baltimore, USA.

Langland, R. and Baker, N. 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201.

Pellerin, S., Laroche, S., Morneau, J. and Tanguay, M. 2006. Estimation of adjoint sensitivity gradients in observation space using dual (PSAS) formulation of the MSC operational 4D-Var. In: *Proceedings of the Seventh International Workshop on Adjoint Applications in Dynamic Meteorology*. 9–13 October, 2006, Obergurgl, Austria.

Tan, D., Andersson, E., Fisher, M. and Isaksen, L. 2007. Observing system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. Tech. Memo. 510, ECMWF.

Trémolet, Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**, 2483–2504.

Trémolet, Y. 2007a. First-order and higher-order approximations of observation impact. *Meteorol. Z.* **16**, 693–694.

Trémolet, Y. 2007b. Incremental 4D-Var convergence study. *Tellus* **59A**, 706–718.

Trémolet, Y. 2007c. Model error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.* **133**, 1267–1280.

Wang, Z., Navon, I. M., Le Dimet, F.-X. and Zou, X. 1992. The second order adjoint analysis: theory and applications. *Meteorol. Atmos. phys.* **50**, 3–20.

Wu, W.-S., Purser, J. and Parrish, D. 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.* **130**, 2905–2916.

Xu, L., Langland, R., Baker, N. and Rosmond, T. 2006. Development and testing of the adjoint of NAVDAS-AR. In: *Proceedings of the Seventh International Workshop on Adjoint Applications in Dynamic Meteorology*, 9–13 October, 2006, Obergurgl, Austria.

Zhu, Y. and Gelaro, R. 2008. Observation sensitivity calculations using the adjoint of the gridpoint statistical interpolation (GSI) analysis system. *Mon. Wea. Rev.* **136**, 335–351.