Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation

By GERALD DESROZIERS* and SERGUEI IVANOV
Météo-France, France

(Received 27 April 2000; revised 27 November 2000)

SUMMARY

Following the a posteriori diagnosis approach proposed by some authors, a practical computation of the expectation of sub-parts of the value of a cost function at the minimum is shown to be feasible by using a randomization technique based on a perturbation of observations or background fields. These computations allow the tuning of observation-error weighting parameters by applying a simple iterative fixed-point procedure. The procedure is first tested in a simplified variational scheme on a circular domain and then in a similar scheme but with the addition of the vertical coordinate. The relationship between the proposed approach and the Generalized Cross Validation is also shown. A test in the French Action de Recherche Petite Echelle Grande Echelle (ARPEGE) three-dimensional variational framework with both simulated observations and background fields is finally performed. It shows that a complete description of observation-error parameters can be retrieved with only a few iterations and, thus, at a reasonable cost.

KEYWORDS: Parameter estimation Variational assimilation

1. INTRODUCTION

Most modern assimilation schemes basically rely on linear estimation theory, or on an extension of this theory. In such an approach, each observation is given a weight that is proportional to the inverse of its specified error variance, measuring the confidence or the precision given to this particular observation. Because of the poor accuracy of certain observations or their sparse density in some areas, practical implementations of operational analysis schemes are based on the use not only of proper observation sets but also of background fields, given by a short-range forecast. In fact, these background fields can be seen as another source of observations (Talagrand 1997) with a given confidence that corresponds to the forecast-error covariances. Because the final analysis is very dependent on the specification of the relative weights given to each source of observations, through the error covariances, and because these errors are not perfectly known, a large potential for improvement on analyses is offered by methods producing a posteriori diagnoses of a mis-specification of a priori errors, or by procedures allowing an adaptive tuning of these parameters.

On the other hand, large operational centres are now using, or have planned to use, assimilation schemes based on a three-dimensional or four-dimensional variational (3D-Var or 4D-Var, respectively) approach, that especially allows the use of a wider range of observations (Lewis and Derber 1985; Courtier and Talagrand 1987). Diagnoses based on statistics of innovations (the differences between observations and background) (Bennett et al. 1993), or of departures between observations (including the background) and the minimizing solution (Talagrand 1999) that can be applied in a variational framework, have been proposed. In particular, it has been shown that a simple diagnosis is the value at the minimum of the cost function that measures the distance between observations and analysis.

In that case, the idea is to perform a posteriori statistics of such a criterion on a sufficiently large number of realizations of the analysis process in order to verify the a priori setting of certain parameters.

* Corresponding author: Météo-France, CNRM/GMME/RECYF, 42 av. G. Coriolis, 31057 Toulouse Cedex, France.
A different kind of method tries to recover an appropriate tuning of error parameters from a single set of observations used in one analysis. In particular, methods based on the maximum-likelihood approach (Dee 1995) have been proposed. A second type of procedure inspired by the Generalized Cross Validation (GCV) approach was also introduced (Wahba and Wendelberger 1980) and proved to be feasible in a variational context (Wahba et al. 1995). Both types of methods have been applied to estimate a limited number of parameters, simultaneously with the estimation of the analysis. In that sense, these methods differ from others retrieving a specification of observation or background errors from a comparison over a period of time between observations and short-range forecasts (Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986).

This paper presents a method based on diagnoses of observations-minus-analysis differences but which aims to perform an adaptive tuning of observation-error parameters from a single batch of observations and background fields.

In section 2 the general variational framework for the implementation of such an approach is presented, as well as the observations-minus-analysis diagnosis approach; a practical way for computing these diagnoses is also described. Section 3 introduces the proposed procedure in a simple case, where one tries to tune the ratio between background and observation terms; the link with GCV for this simple case is also shown; an application in a simplified one-dimensional (1D) case is given. Then, it is shown how the procedure can be generalized to the determination of weights of parts of the background and observation terms, and an application to the tuning of the vertical profile of observation or background errors is performed still with a simple toy variational problem, yet including a second vertical dimension. Section 4 presents an application of the procedure, used to optimize radiosonde errors, in a more realistic framework given by the operational 3D-Var scheme used at Météo-France. Conclusions and perspectives are formulated in section 5.

2. Diagnoses in a Variational Scheme

(a) Variational formulation

The principle of the incremental formulation of 3D/4D-Var algorithms (Courtier et al. 1994) is to seek the increment $\delta x$ to be added to the background $x^b$—so that the analysis is given by $x^a = x^b + \delta x$—which minimizes the cost function

$$J(\delta x) = J^b(\delta x) + J^o(\delta x)$$
$$= \frac{1}{2} \delta x^T B^{-1} \delta x + \frac{1}{2} (d - H \delta x)^T R^{-1} (d - H \delta x).$$

The background term $J^b$ measures the distance between the analysis $x^a$ and the short-range forecast $x^b$, with $B$ the forecast-error covariance matrix. In the observation term $J^o$, $H$ is the linearized observation operator, $R$ stands for the observation-error covariance matrix, including representativeness errors (Lorenc 1986), and $d = y^o - H(x^b)$ is the innovation vector, with $H$ the observation operator that allows the computation of the model equivalents in the space of the observations.

The solution of the minimization of $J$ is given by

$$\delta x^a = K d = K \{y^o - H(x^b)\}, \quad (1)$$

where $K = BH^T (HBH^T + R)^{-1}$ is the gain matrix.
The previous formalism corresponds to the 3D-Var case; however, the extension to the temporal dimension (4D-Var) can be easily introduced: then H includes the integration of the forecast model, $x^a$ stands for the initial conditions at the beginning of the assimilation period, and an outer loop in the minimization process is introduced in order to take into account some parts of the nonlinearities contained in H. Although the following developments stand in a 4D-Var framework, this paper focuses on the 3D-Var case for the sake of simplicity.

(b) Statistical average of the terms in J

Following Talagrand (1997), the complete vector of observations, $z^o$, can be seen as the two-component vector of proper observations, $y^o$, with dimension $p$, and of the background estimate $x^b$, with the same dimension $n$ as the true state $x^i$: $z^o = [(x^b)^T (y^o)^T]^T$ and

$$x^b = x^i + \epsilon^b,$$

where $\epsilon^b$ is the vector of unknown forecast errors with covariance matrix $B$, $y^o = H(x^i) + \epsilon^o$, where $\epsilon^o$ is the vector of unknown observation errors with covariance matrix $R$. Thus, $z^o$ could also be written

$$z^o = \Gamma(x^i) + \epsilon,$$

with $\Gamma$ the complete observation operator and $\epsilon$ the vector of forecast and observation errors, with dimension $n + p$.

An important result pointed out by Talagrand (1999) is that, if $J_j$ stands for a term of $J$, which is the sum of $m_j$ elements, then the expectation of $J_j$ at the minimum is

$$E(J_j(x^a)) = \frac{1}{2} \{m_j - \text{Tr}(\Gamma_j^T S_j^{-1} \Gamma_j P^a)\},$$

where $\Gamma_j$ and $S_j$ are, respectively, the linearized observation operator and the observation-error covariance matrix associated with these $m_j$ elements (a proof of this relation is given in appendix A). Tr is the trace of the matrix to which it is applied. Here, $P^a$ stands for the estimation analysis-error covariance matrix resulting from the analysis with the whole set of $m$ pieces of observations ($m = \sum_j m_j$ and $J = \sum_j J_j$).

A first application of the previous expression is for the complete observation term $J^o$, with, in this case, $\Gamma_j = H$ and $S_j = R$:

$$E(J^o) = \frac{1}{2} \{p - \text{Tr}(H^T R^{-1} H P^a)\} = \frac{1}{2} \{p - \text{Tr}(H^T K^T)\} = \frac{1}{2} \text{Tr}(I_p - HK),$$

where $I_p$ stands for the identity matrix with order $p$, and using the fact that $K = P^a H^T R^{-1}$. Similarly, the application for the complete background term, with $\Gamma_j = I_n$ and $S_j = B$ gives

$$E(J^b) = \frac{1}{2} \{n - \text{Tr}(B^{-1} (B - KHB))\} = \frac{1}{2} \{n - \text{Tr}(I_n - B^{-1} KB)\} = \frac{1}{2} \text{Tr}(KH).$$
where \( I_n \) is the identity matrix with order \( n \) and using the relation \( P^a = B - KHB \).

Thus, since \( \text{Tr}(HK) = \text{Tr}(KH) \), it follows that

\[
E(J) = E(J^a) + E(J^b) = 1/2 \text{Tr}(I_p) = p/2.
\]

This means that if the observation-error covariances are properly specified, then the expectation of the value of the cost function at the minimum is simply proportional to the number of proper observations, \( p \), which also corresponds to the number of degrees of freedom in the analysis problem. As shown by some authors (Bennett et al. 1993; Talagrand 1999), this provides a simple a posteriori diagnosis that requires no extra cost: a deviation of the expectation of \( J \) from \( p/2 \) will indicate a mis-specification of the statistics of the observation errors. It can also be shown that if observation errors are normally distributed then \( J \) has a \( \chi^2 \)-distribution with \( p \) degrees of freedom. This suggests that each part, \( J_j \), of \( J \) should also have a \( \chi^2 \)-distribution but with a rough count of degrees of freedom given by \( m_j - \text{Tr}(\Gamma_j^{-1}S_j^{-1} \Gamma_j P^a) \).

Statistics of parts of \( J \) through Eq. (2) can be powerful diagnosis tools: however, there is a difficulty in their estimation coming from the computation of \( \text{Tr}(\Gamma_j^{-1}S_j^{-1} \Gamma_j P^a) \). The following section introduces a practical way for computing such an expression even when \( m_j \) is large.

(c) Randomized estimation of \( \text{Tr}(KH) \) and \( \text{Tr}(HK) \)

The computation of \( E(J^a) \) is examined first: from the relation given in the previous section, it requires the computation of \( \text{Tr}(KH) \). The idea is to estimate this trace by a randomized method, using a method proposed by Girard (1987) and applied to the GCV computation (Girard 1991; Wahba et al. 1995), which implies the computation of the same expression (see section 3): it can be shown that a randomized estimation \( \text{Rand Tr}(KH) \) of \( \text{Tr}(HK) \) is given by

\[
\text{Rand Tr}(HK) = \text{Rand Tr}(R^{-1/2}HKR^{1/2}) = (R^{-1/2} \xi)\text{T}HKR^{1/2} \xi,
\]

where \( \xi \) is a \( p \)-dimensional vector of numbers with a standard Gaussian distribution (mean zero and variance one).

In practice this is not sufficient, since \( K \) is not explicitly computed in an operational variational scheme. However, Eq. (1) yields

\[
H\delta x^a(y^o+\delta y^o) - H\delta x^a(y^o) \simeq HK\delta y^o,
\]

where \( \delta x^a(y^o+\delta y^o) \) and \( x^a(y^o) \) are the analysis increments obtained with perturbed and unperturbed observations, respectively.

Combining both previous expressions, with \( \delta y^o = R^{1/2} \xi \), a randomized estimate of \( \text{Tr}(HK) \) is given by

\[
\text{Rand Tr}(HK) = (R^{-1/2} \xi)\text{T}(H\delta x^a(y^o+R^{1/2} \xi) - H\delta x^a(y^o)).
\]

The accuracy of this estimate depends not only on the value of \( \text{Tr}(HK) \) itself, but also on the number of observations (Girard 1991; Wahba et al. 1995).

Since \( \text{Tr}(HK) = \text{Tr}(KH) \), this randomized estimation can also be used to compute \( E(J^b) \). However, a similar procedure can be employed to determine \( \text{Tr}(KH) \) separately (this independent computation is also given here to introduce the computation of the expectation of parts of \( J^b \) addressed in a following section): using Eq. (1) again, it
follows that

$$\delta x^b_{(x^b + \delta x^b)} - \delta x^b_{(x^b)} \simeq -KH\delta x^b,$$

and then a randomized estimate of Tr(KH) can also be given by

$$\text{Rand Tr}(KH) = -\left(\mathbf{B}^{-1/2}\xi\right)^\top(\delta x^b_{(x^b + \mathbf{B}^{1/2}\xi)} - \delta x^b_{(x^b)}),$$

where $\xi$ is a vector of random numbers with dimension $n$ in this case. Both previous expressions for Tr(HK) and Tr(KH) provide a practical means to compute $E(\mathbf{J}^\circ)$ and $E(\mathbf{J}^\circ)$: each computation requires two analysis steps, with unperturbed observations ($\mathbf{y}^0$ and $\mathbf{x}^b$, respectively) and perturbed observations ($\mathbf{y}^0 + \mathbf{R}^{1/2}\xi$ and $\mathbf{x}^b + \mathbf{B}^{1/2}\xi$, respectively).

Note that Fisher (1999) proposed a different randomized estimation of Tr(KH) to compute GCV using the relation $\mathbf{K} = \mathbf{P}^\top\mathbf{H}^\top\mathbf{R}^{-1}$ and applying a Lanczos algorithm to the Hessian matrix $J^\circ = (\mathbf{P}^\top)^{-1}$ of the cost function.

3. Adaptive Tuning Procedure

(a) On-line tuning of the ratio between background and observation errors

The problem of tuning the weighting parameters of the background and observation terms is addressed first in this section. If $s^b$ and $s^o$ denote these two parameters, then the cost function introduced in the previous section can be rewritten as

$$J(\delta x) = \frac{1}{s^b} J^b(\delta x) + \frac{1}{s^o} J^o(\delta x).$$

As in some applications of the GCV or in the maximum-likelihood methods, the procedure that is proposed here tries to optimize this couple of parameters, or simply their ratio, $\lambda = s^o / s^b$, from a single batch of observations. The rationale behind this is that, if $s^b$ and $s^o$ are the proper weights to introduce, then the values of $(1/s^b)^2 J^b(\delta x)$ and $(1/s^o)^2 J^o(\delta x)$ should be close to their expected values, that is to say $(1/2) \text{Tr}(KH(s^b, s^o))$ and $(1/2) \text{Tr}(I_p - HK(s^b, s^o))$, respectively.

Thus, defining $s^b = 2J^b/\text{Tr}(KH)$ and $s^o = 2J^o/\text{Tr}(I_p - HK)$, the idea is to find $s^b$ and $s^o$ such that

$$\begin{align*}
s^b &= s^b(s^b, s^o) \\
s^o &= s^o(s^b, s^o). \tag{4}
\end{align*}$$

At this stage, it is important to point out the similarity between this simple approach and GCV: the GCV criterion, which is usually minimized with respect to the ratio $\lambda$, is defined by

$$V(\lambda) = \frac{(\mathbf{d} - H\delta x^b)^\top \mathbf{R}^{-1} (\mathbf{d} - H\delta x^a)}{(\text{Tr}(I_p - HK))^2}$$

$$= \frac{S^o(\delta x^a)}{\text{Tr}(I_p - HK)}. \tag{5}$$

A randomized estimation Rand $V$ of $V(\lambda)$ can also be computed using the randomization procedure previously described.
Therefore, the GCV criterion is the ratio between $S^o$ and $\text{Tr}(I_p - HK)$, this last expression appearing itself in the computation of $S^o$. In appendix B, it is shown that if $s^b$ and $s^o$ are the optimal values of these parameters, then $V$ reaches a minimum for their ratio $\lambda$ and also $\lambda = s^{o2}/s^{b2} = S^o(s^b, s^o)$, so that both approaches are equivalent in this case.

The form of the nonlinear pair of equations (4) with respect to $s^b$ and $s^o$, suggests the use of an iterative fixed-point method with the sequence
\begin{align}
\left\{ s^b_{k+1} = [S^b(s^b_k, s^o_k)]^{1/2} \\
\right.
\left. s^o_{k+1} = [S^o(s^b_k, s^o_k)]^{1/2} \right\} \tag{7}
\end{align}

at iteration $k$.

(b) Test in a simplified one-dimensional variational formulation

A toy problem was constructed to test the iterative tuning procedure introduced in the previous section. The framework is given by an analysis problem on a circular domain (say on an earth meridian). A spectral Fourier decomposition of the signal (assumed here to be a temperature increment) is used. The variational problem is formulated as in Eq. (3), where $\delta x$ stands for the vector of spectral coefficients.

The length of the domain is arbitrarily set to 40 000 km and the truncation to 100, which gives $n = 201$ spectral coefficients. To define the background-error covariance matrix $B$, a Gaussian shape structure function is first defined in physical space with a 300 km length-scale. Assuming homogeneity on the domain, it can be shown that matrix $B$ is diagonal and that its diagonal is obtained by applying the Fourier transform to the Gaussian correlation in physical space. Here, it is first assumed that the square root of the background error is also uniform on the domain and equal to $1$ K, so that matrix $B$ also corresponds to a correlation matrix.

The observation operator $H$ is obtained thanks to an inverse Fourier transform and a simple linear interpolation at the observation location: thus, $H$ is linear in this particular case. We also consider that the $p$ observations have the same error variances and, therefore, matrix $R$ is simply defined as the $p \times p$ identity matrix, the normalization by the variance being introduced through coefficient $s^{o2}$ in Eq. (3). The number of observations, $p$, is set to 100.

In order to precondition the problem, a change of variable is performed with $\delta x = B^{1/2} \chi$, where $\chi$ is the new control variable, so that the cost function to minimize becomes $J(\chi) = 1/2 \chi^T \chi + J^0(B^{1/2} \chi)$.

The minimization is performed with a Quasi-Newton algorithm (Gilbert and Lemaréchal 1989) involving a maximum of 70 iterations. The solution of this simple linear problem is also explicitly computed using Eq. (1).

To test the tuning procedure introduced in the previous section, both background and observation errors are simulated. If $x^i$ denotes the true signal, then $x^i = x^b - \epsilon^b$. We will further simplify the problem by assuming that $x^b = 0$ (so that $x^i = -\epsilon^b$). If $\zeta^b$ is an $n$-dimensional vector of random numbers with a Gaussian distribution and variance 1, then writing $\epsilon^b = B^{1/2} \zeta^b$ is a way of simulating background errors consistent with the matrix $B$ since $E(\epsilon^b \epsilon^b^T) = B^{1/2} E(\zeta^b \zeta^b^T) B^{1/2} = B$ (using the fact that $E(\zeta^b \zeta^b^T)$ is the identity matrix since $\zeta^b$ is a vector of random values with variance 1).

Similarly, a $p$-dimensional vector of observation errors can be obtained with $y^o = H(x^i) + R^{1/2} \zeta^o$, where $\zeta^o$ is a vector of random numbers with a Gaussian distribution. Figure 1 shows the result of such a definition of $x^i$ and $y^o$; here $s^o = 0.1$. 

...
TUNING OF OBSERVATION-ERROR PARAMETERS

Figure 1. True signal $x^t$ (full line) and observations $y^o$ (circles) simulated with the procedure described in section 3(b); the length-scale of the structure function is 300 km and $\lambda = s^o / s^b = 0.01$. See text for further details.

Figure 2. Exact criterion $S^0$ (full line with circles) as a function of $s^o$, computed with the explicit solution, and randomized estimation of $S^0$ (dashed line with circles) computed with the two solutions $x^s_{(r)}$ and $x^s_{(r')}$ resulting from the minimization of the cost function $J$ with the Quasi-Newton descent algorithm; solid line with stars and dashed line with stars: same as previously but for the criterion $S^b$. See text for further details.
The first step of this test addresses the problem of recovering the correct value of \( s^2 \) or \( \lambda = s^2 / s_b^2 \) since here, the parameter \( s_b^2 \) is assumed to be known (\( s_b^2 = 1 \)). Figure 2 shows the variation of \( S^0 \) with respect to \( s^2 \), with two computations: one with the explicit solution \( x^a \) and the exact computation of \( \text{Tr}(\mathbf{I}_P - \mathbf{K} \mathbf{H}) \) and another evaluation of \( S^0 \) obtained with the pair of solutions \( x_a^{(s^0)} \) and \( x_{b+R}^{(s^0)} \), produced by minimizing the cost function \( J \) with the Quasi-Newton descent algorithm. Accordingly, with the approach introduced in section 3(a), it appears that both exact and randomized computations of \( S^0 \) are very similar, and that the relation \( s^2 = S^0 \) is fulfilled for a value of \( s^2 \) very close to the true value (0.01) used for simulating the observations. In this case, both exact and randomized computations of \( S^0 \) show that the relation \( s_b^2 = S^0 \) is true for the optimal value \( s_b^2 = 1 \) used for simulating the vector of background errors (Fig. 2), which is again in agreement with the pair of equations (4).

Then, the simple iterative method proposed in section 3(a) with Eq. (7) is applied, but for the determination of \( s^0 \) only (\( s_b^2 \) is kept equal to the optimal value). Table 1 shows that the optimal value is retrieved with this algorithm, with only a few iterations, whatever the initial value of the parameter is (no modification of \( s^0 \) appears after these few iterations and, therefore, Table 1 displays results until iteration 5 only). The value of the GCV criterion \( V \) is also reduced in agreement with Fig. 3 and with the theoretical justification given in appendix B.
TUNING OF OBSERVATION-ERROR PARAMETERS

| TABLE 1. VALUES OF OBSERVATION-ERROR WEIGHTING PARAMETERS ($s^o$) ALONG THE FIXED-POINT ITERATIONS |
|--------------------------------------------------|---------------------------------|----------------------|
| $s^o$ it. 0                                      | 0.01                            | 0.10                 |
| r.m.s.                                           | 0.1                             | 0.08                 |
| $V$                                              | $0.4 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |
| $s^o$ it. 1                                      | 0.06                            | 0.09                 |
| r.m.s.                                           | 0.08                            | 0.09                 |
| $V$                                              | $0.1 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |
| $s^o$ it. 2                                      | 0.08                            | 0.09                 |
| r.m.s.                                           | 0.08                            | 0.08                 |
| $V$                                              | $0.1 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |
| $s^o$ it. 3                                      | 0.09                            | 0.10                 |
| r.m.s.                                           | 0.09                            | 0.09                 |
| $V$                                              | $0.1 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |
| $s^o$ it. 4                                      | 0.09                            | 0.10                 |
| r.m.s.                                           | 0.09                            | 0.09                 |
| $V$                                              | $0.1 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |
| $s^o$ it. 5                                      | 0.09                            | 0.10                 |
| r.m.s.                                           | 0.09                            | 0.09                 |
| $V$                                              | $0.1 \times 10^{-3}$            | $0.1 \times 10^{-3}$ |

The three right-hand columns of the table correspond to three different starting points; the values of the root mean square (r.m.s.) error and of the Generalized Cross Validation (GCV) criterion ($V$) are also given at the beginning and at the end of the iterations (it. 1–it. 5); the last three rows give the reference value of $s^o$ used for the simulation of observations ($s^o)^y$ and the corresponding r.m.s. error and GCV criterion obtained with the use of this value in the analysis.

Indeed, this first test problem is rather simple due to the choice of homogeneous error covariances: the accuracy of such an estimation of the parameter with more realistic forms of covariances will have to be further investigated.

(c) Tuning the weights given to parts of the background or observation terms

In the previous section, it was shown that a simple iterative algorithm can be used for retrieving the global ratio between the background and the observation terms in a variational formulation. The possibility of tuning the weights given to parts of these two terms is now investigated.

First, if $J^o_j$ stands for a part of $J^o$ with $p_j$ observations, and using again Eq. (2) introduced in section 2(b), then the following sequence of equalities can be written:

$$
E(J^o_j) = \frac{1}{2} \{ p_j - \text{Tr}(H_j^T R_j^{-1} H_j P^a)) \\
= \frac{1}{2} \{ p_j - \text{Tr}(H_j^T R^{-1} P_j H P^a)) \\
= \frac{1}{2} \{ p_j - \text{Tr}(P_j H P^a H^T R^{-1} P_j) \\
= \frac{1}{2} \{ p_j - \text{Tr}(P_j (H K) P_j^T)),
$$

where $P_j$ is the projection operator that allows the passage from the complete set of observations to the subset of $p_j$ observations, and $R_j$ is the observation-error covariance matrix associated with this subset.

This means that once matrix $HK$ is known, then the expectation of a part of $J^o$ can be computed by applying the proper projection operator.
Similarly, the expectation of a part, $J_j^b$, of the background term can be computed using the following relations:

$$E(J_j^b) = \frac{1}{2} \{ n_j - \text{Tr}(P_j^T B_j^{-1} P_j P^a) \}$$

$$= \frac{1}{2} \{ n_j - \text{Tr}(B_j^{-1} P_j^T P_j P^a) \}$$

$$= \frac{1}{2} \{ n_j - \text{Tr}(P_j P^a B_j^{-1} P_j) \}$$

$$= \frac{1}{2} \{ n_j - \text{Tr}[P_j (I_n - (KH)) P_j^T] \},$$

where $P_j$ is now the projection operator that allows the passage from the complete background vector $x^b$ to a subset with $n_j$ elements, and $B_j$ is the background-error covariance matrix associated with this subset.

Once more, these relations do not suffice, since neither $HK$ nor $KH$ are explicitly computed in a variational framework. A randomized estimation of $E(J_j^o)$ and $E(J_j^o)$, can however, be proposed in the form of

$$\text{Rand} \text{Tr}[P_j (HK) P_j^T] = (P_j R^{-1/2} \xi)^T P_j (H \delta x^a_{(y^o + R^{1/2} \xi)} - H \delta x^a_{(y^o)}), \quad (8)$$

and

$$\text{Rand} \text{Tr}[P_j (KH) P_j^T] = -(P_j B^{-1/2} \xi)^T P_j (\delta x^a_{(x^b + B^{1/2} \xi)} - \delta x^a_{(x^b)}). \quad (9)$$

Equation (8) allows the computation of the expectation of parts of $J^o$, from only two analyses, one with the complete set of true values $y^o$ of the observations, and the other with the whole set of perturbed observations, $y^o + R^{1/2} \xi$, followed by the application of $P_j$ to the observation perturbation set $R^{1/2} \xi$ and to the difference between the pair of perturbed and unperturbed analyses.

Similarly, Eq. (9) proves that the expectation of parts of $J^b$, can be determined from two other analyses, one with the true values $x^b$ of the background field, and the other one with a perturbed background $x^b + B^{1/2} \xi$, and then by using the corresponding projection operator $P_j$.

The cost function defined in Eq. (3) can again be rewritten with

$$J(\delta x) = \sum_j \frac{1}{s^b_j} J^b_j(\delta x) + \sum_j \frac{1}{s^o_j} J^o_j(\delta x), \quad (10)$$

where $s^b_j$ and $s^o_j$ are, respectively, the background and observation-error weighting parameters supposed to be homogeneous for a given subset of observations, $j$.

Thus, following the same idea as in section 3(a), and defining $S^b_j = 2J^b_j/(n_j - \text{Tr}[P_j (I_n - (KH)) P_j^T])$ and $S^o_j = 2J^o_j/[P_j - \text{Tr}[P_j (HK) P_j^T]]$, a procedure for finding the optimal values of parameters $s^b_j$ and $s^o_j$ is to solve the following set of equations

$$\begin{cases} s^b_j = S^b_j \\ s^o_j = S^o_j. \end{cases} \quad (11)$$

(d) Test in a simplified two-dimensional variational formulation

The possibility of recovering error variances with such a procedure is first tested in a simple problem similar to the previous case, but with the addition of the vertical
TUNING OF OBSERVATION-ERROR PARAMETERS

Figure 4. Two-dimensional simulated true signal $x'$; the length-scale of the structure function is $\mathcal{L}_h = 300$ km in the horizontal and $\mathcal{L}_v = 1/200 \mathcal{L}_h$ in the vertical; isotherms are every 1 K; circles show the location of the observations. See text for further details.

coordinate. Such a framework allows the opportunity to specify different background- or observation-error variances at each level. The same spectral resolution as in the 1D case is used, but now with six vertical levels from 0 to 10,000 m. If $m$ stands for the number of spectral coefficients in the horizontal and $l$ is the number of levels, then $\mathbf{B}$ is defined as a block-diagonal matrix, each $l \times l$ block corresponding to the background correlations for a given horizontal wave number. Although such a formulation allows the full specification of non-separable correlation structure functions with a different vertical correlation for each wave number (Rabier et al. 1998), the correlations are simply assumed to be all the same here, and defined by a Gaussian function with length-scale $\mathcal{L}_v = 1/200 \mathcal{L}_h$, where $\mathcal{L}_h$ is the length-scale of the horizontal correlation (300 km as in the 1D case). As in the 1D case, a fictitious true signal $x'$ is simulated by $x' = -\mathbf{\epsilon}^b = -\mathbf{B}^{1/2} \mathbf{\xi}^b$, where $\mathbf{\xi}^b$ is a vector of $m \times l$ random values with a Gaussian distribution.

- Firstly, the background-error weighting parameters, $s^b_j$, are assumed to be equal to 1 at each level. Figure 4 shows the field obtained with such a constant background-error variance. On the contrary, the vector of observations is simulated with $\mathbf{y}^o = \mathbf{H}(x') + \mathbf{R}^{1/2} \mathbf{\xi}^o$, where $\mathbf{\xi}^o$ is a vector of random numbers with a Gaussian distribution, but with $\mathbf{R}$ containing values of the standard deviations of observation errors that increase with altitude: $(s^b_j, j = 1, l) = (1.0, 1.5, 2.0, 2.5, 3.0, 3.5)$ (here, the observations are placed at field levels for simplicity). The values of the background-error parameters, $s^b_j$, being kept constant, the optimal values of the observation-error parameters, $s^o_j$, are searched for by the same iterative fixed-point procedure as in the 1D case, but for all $l$ levels simultaneously: $(s^o_{j,k+1} = S^o_{j,k}, j = 1, l)$, where $k$ stands for the iteration index. The results of such an iterative procedure are presented in Table 2: they show that a vertical profile of observation-error weighting parameters can be obtained after a few
The first row gives the values at the initial point of the iterative procedure (it. 0) (constant values in this case) and the second row shows the corresponding root mean square (r.m.s.) errors of the analyses performed with these values; the r.m.s. error at iteration 5 (it. 5) is also given; the last two rows give the reference values of $s_j^0$ used for the simulation of observations and the corresponding r.m.s. errors obtained with the use of these values in the analyses.

**Table 3.**

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>Level 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_j^0$ it. 0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>0.52</td>
<td>0.52</td>
<td>0.59</td>
<td>0.60</td>
<td>0.74</td>
</tr>
<tr>
<td>$s_j^0$ it. 1</td>
<td>1.36</td>
<td>1.61</td>
<td>1.93</td>
<td>2.19</td>
<td>2.32</td>
</tr>
<tr>
<td>2</td>
<td>1.35</td>
<td>1.65</td>
<td>2.11</td>
<td>2.47</td>
<td>2.65</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>1.69</td>
<td>2.14</td>
<td>2.50</td>
<td>2.69</td>
</tr>
<tr>
<td>4</td>
<td>1.33</td>
<td>1.69</td>
<td>2.15</td>
<td>2.50</td>
<td>2.69</td>
</tr>
<tr>
<td>5</td>
<td>1.33</td>
<td>1.69</td>
<td>2.15</td>
<td>2.50</td>
<td>2.69</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>0.47</td>
<td>0.45</td>
<td>0.41</td>
<td>0.41</td>
<td>0.45</td>
</tr>
<tr>
<td>$(s_j^0)^*$</td>
<td>1.00</td>
<td>1.50</td>
<td>2.00</td>
<td>2.50</td>
<td>3.00</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>0.46</td>
<td>0.45</td>
<td>0.41</td>
<td>0.41</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Same as Table 2, but for the background-error weighting parameters $(s_j^b)$ (the values of $s_j^0$ are kept constant during the iterations).

iterations only. Note that, as in the 1D case, each iteration requires two minimizations of the cost function $J$ with unperturbed and perturbed observations.

- Secondly, the background-error weighting parameters, $s_j^b$, are now specified as varying in the vertical $(s_j^b, j = 1, l) = (1.0, 1.5, 2.0, 2.5, 3.0, 3.5)$ in the background-error simulations and, on the contrary, the observation-error weighting parameters, $s_j^0$, are kept constant with a value of 1 in the simulation of observations. This time, the iterative fixed-point procedure is used for optimizing the $s_j^b$ profile but the $s_j^0$ profile is unchanged during the iterations: each iteration now requires two computations of the analysis with unperturbed and perturbed backgrounds. Table 3 shows that, again, this iterative method allows the recovery of a profile that is very close to the optimal one: although the values are not exactly the same as those used in the simulation of the background errors, the corresponding analysis errors are very close.
• Finally, the values of $s_j^b$ and $s_j^o$ used for background and observation errors are kept the same as in the previous case, but now a simultaneous iterative tuning of both sets of error parameters is performed: in this case each fixed-point iteration requires a pair of minimizations of $J$ with unperturbed and perturbed observations for computing the $S_{j,k}^i$ expressions at iteration $k$, and another couple of minimizations with unperturbed and perturbed backgrounds to compute the $S_{j,k}^o$ expressions. Again, it can be checked (Table 4) that this iterative procedure allows the production of a very good approximation of the optimal values of the parameters, although the convergence is slightly slower here for the determination of the $s_j^b$ weights only.

Again, these first good results are certainly related not only to the simplicity of the assumptions made regarding the form of the error covariances, but also to the vertical length-scale of the correlation that is short enough to leave uncoupled the information coming from the different levels; some tests have shown that the estimation becomes less accurate when this length-scale increases. This means that more work is needed to investigate the scope of validity of the procedure in a more realistic framework. Such a preliminary test is described in next section, but for the tuning of the $s_j^o$ parameters only.

4. APPLICATION IN A THREE-DIMENSIONAL VARIATIONAL FRAMEWORK

A test of the parameter estimation procedure is performed here in the framework provided by the French Action de Recherche Petite Echelle Grande Echelle (ARPEGE) 3D-Var assimilation scheme. This scheme is based on an incremental formulation (Thépaut et al. 1998), with a global spectral model (Courtier et al. 1991). Although a particular feature of the ARPEGE system is the use of a stretched geometry that operationally provides a locally very high resolution for the trajectory, the spectral
resolutions for both the trajectory and the increments are set identical (T63 with no stretching), in order to simplify the framework of this test.

Once more, the idea is to simulate both background and observations corresponding to a given set of background and observation errors and then to recover the optimal specification of the variances in the analysis step, that should also give the best analysis. It is assumed that the true values $x^t$ of the ARPEGE variables (temperature, vorticity, divergence, and specific humidity at 31 levels, and surface pressure) are given by a particular 6 h forecast (valid at 12 UTC 17 February 1997 here). Then, a vector $\epsilon^b$ of forecast errors is obtained by applying the square root of the background-error covariance matrix $B^{1/2}$ to a vector of random numbers (this can be done in practice by applying the change-of-variable operator, $L = B^{1/2}$, introduced for pre-conditioning the problem as in section 3(b)). Given that the $J^b$ term formulation is multivariate in the ARPEGE 3D-Var method, following the approach developed by Bouttier et al. (1997), this operation produces a vector $\epsilon^b$ in which the temperature and the vorticity fields are, for example, correlated. Figure 5 presents the corresponding simulated background-error field for the temperature at 500 hPa. The simulated observations are given by a set of radio-sounding measurements with true locations (see Fig. 5): the observation errors are produced by using the operational error variances for geopotential, temperature, wind, and specific humidity (relative-humidity measurements are converted to specific-humidity observations before being used in the ARPEGE 3D-Var model). The variation of the standard deviation for these parameters is given by the full lines in Figs. 6 to 9. If $\epsilon^o$ stands for the vector of observation errors, then the observations $y^o$ are produced

Figure 5. Field of temperature errors at 500 hPa produced by the procedure described in section 4; isotherms are every 0.5 K; triangles indicate the location of radio-sounding measurements.
Figure 6. Geopotential-height error profiles used for the simulation of observations (full line), imposed at the beginning of the iterative procedure (dotted line), after 1 iteration (dash-dot line) and after five iterations (dashed line).

Figure 7. Same as Fig. 6 but for temperature.

with \( y^0 = H(x^t) + \epsilon^0 \), with \( x^t \) the fictitious true representation of the atmospheric flow given by the ARPEGE forecast, and \( H \) is the observation operator associated with the radio-sounding observations.

The procedure described in previous sections was implemented in the ARPEGE 3D-Var model to optimize the values of the error variances associated with radiosondes. In this test, no simultaneous search of background-error variances is performed: their values are kept constant along the iterations, and set equal to the optimal values used
Figure 8. Same as Fig. 6 but for wind speed.

Figure 9. Same as Fig. 6 but for specific humidity.

for the background errors. Thus, each iteration of the tuning procedure requires a pair of 3D-Var analyses with unperturbed and perturbed observations. At each iteration, the r.m.s. error between the analysis obtained with the corresponding set of parameters and the fictitious truth is computed for the same variables as in the measurements; in order to identify the impact of tuning the parameters more clearly, this r.m.s. error is shown for a sub-domain roughly corresponding to North America, where the radiosounding coverage is the most homogeneous over the globe (see Fig. 5). The initial vertical profiles of observation-error parameters (dotted lines in Figs. 6 to 9) were
TUNING OF OBSERVATION-ERROR PARAMETERS

Figure 10. Wind-speed root mean square (r.m.s.) errors over a domain covering North America for: the background (full line and crosses), the analysis with wrong initial error profile (dotted line and circles), the analysis with true error profile (full line and circles), and the analysis with optimized profile at iteration five (dashed line and circles); the last two curves are nearly superimposed.

deliberately chosen to be very different from the true ones: they are constant in the vertical for geopotential, temperature and wind speed; the initial vertical profile for specific humidity is obtained by imposing a constant profile for relative-humidity and temperature errors and, then, going to the specific-humidity errors (that explains why the initial profile is not constant in the vertical for specific humidity).

Figures 6 to 9 show that the convergence of the iterative algorithm is extremely fast: the profiles obtained after a single iteration (dash-dot lines) are already very close to those obtained at iteration 5 (dashed lines), although this convergence is slightly slower for the wind speed. Moreover, the figures also show that these optimized profiles are very similar to the operational ones used for the simulation of observations (full lines). Finally, it can be verified that the optimized profiles give an r.m.s. error that almost shows no difference relative to the r.m.s. error resulting from an analysis obtained with the reference parameters (see Fig. 10 for the wind speed r.m.s. error; r.m.s. errors for other parameters are not presented here but show the same behaviour).

5. CONCLUSION

In this paper it was shown that following the a posteriori diagnosis approach proposed by some authors, a practical computation of the expectation of sub-parts of the value of a cost function at the minimum can be achieved by using a randomization technique based on a perturbation of either observations or background fields. Furthermore, these computations allow the tuning of observation-error parameters by applying a simple iterative fixed-point procedure. This procedure was successfully tested in two simplified cases: specifically, the possibility of tuning both observation- and background-error parameters was demonstrated. The link with GCV in the simple problem of determining the ratio between observation and background terms was also pointed out. Finally, a test
in the ARPEGE 3D-Var framework showed that a complete description of observation-error parameters can be retrieved with only a few iterations and, thus, at a reasonable cost (this cost is of the order of some 3D-Var analyses). Although feasible, the determination of background-error parameters in the ARPEGE 3D-Var model was not attempted here, since it is more difficult to implement.

Such promising results with simulated observations and background fields suggest that the procedure should now be applied, in a second step, to real observation sets (however, more tests with such an idealized problem are also planned in order to have a better idea of the accuracy and the sensitivity of the parameter estimates to the number of observations and the background-error covariances). The computation of partial diagnoses for subsets of observations could be, for example, applied occasionally to verify or to adjust the specification of observation errors for particular subsets, as was done here for radio soundings. In particular, the determination of satellite observation-error statistics, such as radiance errors, is of great interest since these errors, including the part associated with the observation operator, are not perfectly known. Another point of interest is the capability of using such a method for tuning the level of representativeness error due to the resolution of the model or of the analysis increment, or to the observation operator, as just mentioned for the use of radiances.

Because background-error covariances should evolve in time, as the observation error statistics are less dependent on the temporal dimension, there should still be a greater advantage to using the proposed approach for performing an adaptive determination of background errors with, for example, a different specification for predefined sub-domains. This was shown to be feasible in the simple two-dimensional variational analysis problem, and the application to the 3D- or 4D-Var ARPEGE analysis scheme will be addressed in future work.

Finally, although this study focused on the optimization of the variance of observation errors, the possibility of adapting the procedure to the determination of correlation-error parameters, such as observation- or background-error length-scales, will also be investigated.

ACKNOWLEDGEMENTS

The authors would like to thank Thierry Bergot, Gwenaelle Hello, Alain Joly, Philippe Lopez and Florence Rabier for helpful discussion and comments on this paper. We also wish to thank the two referees for their constructive comments on the first version of this paper.

APPENDIX A

Proof of Eq. (2)

For any part of $J$, the following sequence of equalities can be written:

$$E(J_j) = \frac{1}{2}E((\Gamma_j x^a - y_j^o)^T R_j^{-1} (\Gamma_j x^a - y_j^o))$$

$$= \frac{1}{2}E(\text{Tr}(R_j^{-1} (\Gamma_j x^a - y_j^o)(\Gamma_j x^a - y_j^o)^T))$$

$$= \frac{1}{2}E(\text{Tr}(R_j^{-1} (\Gamma_j e^a - e_j^o)(\Gamma_j e^a - e_j^o)^T))$$

$$= \frac{1}{2}\text{Tr}(R_j^{-1} [\Gamma_j E(e^a e_j^o^T) + E(e_j^o e_j^o)^T - E((\Gamma_j x^a) e_j^o^T) - E(e_j^o (\Gamma_j x^a)^T)]),$$

since $\Gamma_j x^a - y_j^o = \Gamma_j x^a + \Gamma_j x^T - \Gamma_j x^T - y_j^o = \Gamma_j e^a - e_j^o$, and using the fact that $E(\text{Tr}(.) ) = \text{Tr}(E( . ))$ and the linearity properties of the expectation operator $E$. 
On the other hand, Eq. (1) yields $x^a - x^l = x^b - x^l + K[y^o - H(x^b)]$ or $\epsilon^a = \epsilon^b + K[H(x^b) + \epsilon^o - H(x^b)]$. Then $E(\epsilon^a \epsilon^o^T) = KE(\epsilon^o \epsilon^a^T) = P^aH^TR^{-1}R = P^aH^T$. This last equality gives $E((\Gamma_j x^a)\epsilon^o_j^T) = \Gamma_j E(x^a \epsilon^o_j^T)P_j = \Gamma_j P^a \Gamma_j^T$, where $P_j$ is the projection operator such that $\epsilon^a_j = \Gamma_j \epsilon^o$ (as also introduced in section 3(c)). It is also easy to check that $\text{Tr}[E(\epsilon^o_j (\Gamma_j x^a)^j T)] = \text{Tr}[E((\Gamma_j x^a)\epsilon^o_j^T)] = \text{Tr}(\Gamma_j P^a \Gamma_j^T)$. Since $\Gamma_j E(\epsilon^a \epsilon^o_j^T) \Gamma_j^T = \Gamma_j P^a \Gamma_j^T$ and $E(\epsilon^o_j \epsilon^o_j^T) = R_j$, all those relations yield

$$E(J_j) = \frac{1}{2} \text{Tr}(R_j^{-1}(R_j - \Gamma_j P^a \Gamma_j^T))$$

$$= \frac{1}{2}[m_j - \text{Tr}(R_j^{-1} \Gamma_j P^a \Gamma_j^T)]$$

$$= \frac{1}{2}[m_j - \text{Tr}(\Gamma_j^T R_j^{-1} \Gamma_j P^a)].$$

### APPENDIX B

**Relationship with Generalized Cross Validation minimum**

Equation (5) in section 3(a) can also be written

$$V(\lambda) = \frac{\text{Tr}(I_p - HK)^T R^{-1} (I_p - HK) dd^T}{[\text{Tr}(I_p - HK)]^2}$$

$$= \frac{\text{Tr}(R^{-1}(I_p - HBH^T(HBH^T + \lambda R)^{-1}) dd^T)}{[\text{Tr}(I_p - HBH^T(HBH^T + \lambda R)^{-1})]^2}$$

$$= \frac{\text{Tr}(RD^{-2} dd^T)}{[\text{Tr}(RD^{-1})]^2},$$

with $D = HBH^T + \lambda R$ and $\lambda = s^{o2}/s^{b2}$ as defined in section 3(a).

Computing the derivative of this expression with respect to the parameter $\lambda$, the numerator $N(V')$ of this derivative is

$$N(V') = (-2R^T RD^{-3} dd^T) [\text{Tr}(RD^{-1})]^2 - [\text{Tr}(RD^{-2} dd^T)] [2 \text{Tr}(RD^{-1}) R^T RD^{-2}].$$

This last expression vanishes if $dd^T = D = HBH^T + \lambda R$, that is to say, if the value of $\lambda$ agrees with the statistics $dd^T$ of innovations computed with a single batch of observations and background.

On the other hand, the expression $S^o$ also defined in section 3(a) can be rewritten

$$S^o = \frac{\text{Tr}(\lambda^2 RD^{-2} dd^T)}{\text{Tr}(\lambda RD^{-1})} = \frac{\lambda \text{Tr}(RD^{-2} dd^T)}{\text{Tr}(RD^{-1})},$$

so that $S^o = \lambda = s^{o2}/s^{b2}$ if $dd^T = D$.

Thus, searching for $\lambda$, such that $\lambda = S^o(\lambda)$ or as the minimum of $V(\lambda)$, are two equivalent approaches in this simple case where one simply aims to optimize the ratio between background and observation terms.

### REFERENCES


1452


Fisher, M. 1999 'Implementing GCV in the IPS'. Technical Report, ECMWF, Reading, UK


Lonnberg, P. and Hollingsworth, A. 1986 The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. Tellus, 38, 137–161


G. DESROZIERS and S. IVANOV