



Diagnostics on the cost-function in variational assimilations for meteorological models

Y. Michel

Météo-France and CNRS, CNRM-GAME, UMR3589, Toulouse, France

Correspondence to: Y. Michel (yann.michel@meteo.fr)

Received: 10 July 2013 – Revised: 5 December 2013 – Accepted: 5 December 2013 – Published: 5 February 2014

Abstract. Several consistency diagnostics have been proposed to evaluate variational assimilation schemes. The “Bennett-Talagrand” criterion in particular shows that the cost-function at the minimum should be close to half the number of assimilated observations when statistics are correctly specified. It has been further shown that sub-parts of the cost function also had statistical expectations that could be expressed as traces of large matrices, and that this could be exploited for variance tuning and hypothesis testing.

The aim of this work is to extend those results using standard theory of quadratic forms in random variables. The first step is to express the sub-parts of the cost function as quadratic forms in the innovation vector. Then, it is possible to derive expressions for the statistical expectations, variances and cross-covariances (whether the statistics are correctly specified or not). As a consequence it is proven in particular that, in a perfect system, the values of the background and observation parts of the cost function at the minimum are positively correlated. These results are illustrated in a simplified variational scheme in a one-dimensional context.

These expressions involve the computation of the trace of large matrices that are generally unavailable in variational formulations of the assimilation problem. It is shown that the randomization algorithm proposed in the literature can be extended to cover these computations, yet at the price of additional minimizations. This is shown to provide estimations of background and observation errors that improve forecasts of the operational ARPEGE model.

1 Introduction

Most operational data assimilation schemes in meteorology are loosely based on statistical linear estimation (e.g. Talagrand, 2010). The variational formulation (Le Dimet and

Talagrand, 1986; Courtier et al., 1998; Rabier et al., 2000) has proven very effective in assimilating directly and massively satellite observations, which has been a source of huge progress in numerical weather prediction (Simmons and Hollingsworth, 2002; Rabier, 2005). The cost function to be minimized measures the distances of the atmospheric state to the observations and to the background, weighted by the inverse of their error covariances. These error statistics are unfortunately not very well known. Common methods to estimate observation errors include the innovation approach (Hollingsworth and Lonnberg, 1986; Daley, 1993), where observation error statistics are taken to be spatially uncorrelated and are deduced from the innovation covariance. Background error statistics are estimated within an existing data assimilation scheme, for instance using forecasts at different ranges valid at the same time (Parrish and Derber, 1992; Derber and Bouttier, 1999) or, more recently and inspired by the Ensemble Kalman Filter (e.g. Evensen, 2003), ensembles of perturbed variational assimilations (Fisher, 2003; Kucukkaraca and Fisher, 2006; Belo-Pereira and Berre, 2006).

Despite this recent progress, it is still necessary and useful to check for the consistency between the prescribed statistics and the a posteriori error statistics. Any inconsistency could help to point out imperfections in the specification of the statistics. Objective assessment of the quality of assimilation schemes can only be performed against independent observations that are not used in the estimation process (Talagrand, 1999). Yet internal consistency diagnostics can also prove useful. Two applications of these kinds of diagnostics have been proposed: hypothesis testing (Muccino et al., 2004) and variance tuning (Desroziers and Ivanov, 2001; Chapnik et al., 2006).

Hypothesis testing formally checks whether the obtained solution is consistent with the hypotheses made concerning

the errors in the innovation vector. In the perfect case, the minimum value of the cost function has a χ_p^2 distribution where p is the number of scalar observations assimilated in the system (Bennett et al., 1993; Talagrand, 1999). However, the values of the cost function were found to differ significantly from their expectation in a series of data assimilation experiments with an ocean model (Bennett et al., 2000). Muccino et al. (2004) also studied how various errors in the specified error covariance matrices were affecting the distribution of the value of the cost function at the minimum. A Kolmogorov–Smirnov test was used to identify departures from the χ^2 distribution, with significant skill in some cases.

When certain sets of observations are taken to have uncorrelated errors, it is possible to split the observation term further as the sum of sub-functions. The expectation of these sub-functions can still be expressed (Talagrand, 1999) as the trace of large matrices. Desroziers and Ivanov (2001) (hereafter DI01) proposed to estimate these traces using a randomization algorithm. They also proposed an iterative approach for tuning the variances. It enforces consistency between the observed value of the sub-parts of the cost functions and their theoretical expectations as computed by the randomized trace estimation. This method was later shown (Chapnik et al., 2004) to be similar to a maximum-likelihood approach. Operational implementation of the scheme in the global model ARPEGE¹ from Météo-France was achieved by Chapnik et al. (2006), and the tuning of observation error variances yielded a positive impact on both analyses and forecasts. The implementation in a limited area model was discussed by Sadiki and Fischer (2005). They proposed in particular the use of time or space averages to improve sampling. Finally, the ensemble of variational assimilations with perturbed observations as implemented at Météo-France and ECMWF (Kucukkaraca and Fisher, 2006) typically uses assimilations with perturbed backgrounds and observations. Desroziers et al. (2009) then showed that the previous statistical expressions were a direct by-product of such an ensemble.

More precisely, Dee (1995) and Dee and da Silva (1999) have introduced the maximum-likelihood approach for the estimation of covariance parameters in meteorology. The likelihood function to be optimized with respect to these covariance parameters is the sum of a quadratic form in the innovation vector and the log-determinant of a precision (inverse covariance) matrix. Because of the very large size of the innovation vector (currently of the order of 10^6 elements in the ARPEGE 4D-Var), a direct computation (using for instance Choleski decomposition) is not feasible. The approach that was favored in meteorology is to evaluate the derivatives (rather than the absolute values) of the likelihood function and then using stochastic trace estimation techniques, as demonstrated by Purser and Parrish (2003) in an ideal-

ized context. This stochastic approximation is nearly optimal in a well-define sense as proven by Stein et al. (2013). It is also possible that direct evaluation of the log-likelihood may become feasible even in high dimensions as shown by Aune et al. (2012). This required in particular (i) more efficient evaluation of the trace with probing vectors (Bekas et al., 2007) and (ii) Krylov subspace methods for matrix inversions (Saad, 2003). These approaches look very promising, but as recognized by Purser and Parrish (2003), it is still necessary to experiment with various aspects in meteorology including non-linear observation operators and introduction of a prior term (to stabilize the solution when the number of estimated parameters become large). This study rather builds upon the simpler but convenient approach taken in DI01 for variance tuning. However, it should be stressed that this approach is unlikely to be easily extended to the estimation of other parameters such as localisation lengthscales (Houtekamer and Mitchell, 2001), or to the heteromorphic case where there is a need to compare log-likelihoods associated with covariances that use completely different sets of parameters (Purser and Parrish, 2003).

The aim of this paper is to present a different way of deriving the Bennett–Talagrand criterion. The sub-parts of the cost function at the minimum are expressed as quadratic forms in the innovation vector. This allows the application of the theory of quadratic forms in random variables as summarized in Mathai and Provost (1992) (hereafter M92). The statistical expectation, variances and cross-covariances between sub-parts of the cost-function at the minimum are shown to be related to the trace of matrices in the assimilation. This relies on the assumption that the innovation vector is normally distributed, but some extensions to the Non-Gaussian case have also been derived for some specific distributions. The derivation of these expressions is outlined in Sect. 2. Practical estimation of these terms based on the randomized trace algorithm is demonstrated in Sect. 3. The use of these formulas for variance tuning in the operational ARPEGE model is illustrated in Sect. 4.

2 Statistical distribution of the values of the terms in the cost-function at the minimum

2.1 Variational framework

The best estimate of the atmospheric state, \mathbf{x}_a , is obtained as the solution of the minimization of the cost function:

$$\begin{aligned} J(\mathbf{x}) &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \\ &= \frac{1}{2} \{ (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \\ &\quad + [\mathbf{y}_o - \mathcal{H}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}_o - \mathcal{H}(\mathbf{x})] \} \end{aligned} \quad (1)$$

where any time index has been dropped for convenience. $J_b(\mathbf{x})$ and $J_o(\mathbf{x})$ stand for the background and observation

¹“Action de Recherche Petite Echelle Grande Echelle”, see Pailleux et al., 2000 for a historical description

terms of the cost-function respectively. \mathbf{B} and \mathbf{R} are the background and observation error covariance matrices, and \mathcal{H} is the non-linear observation operator including model time integration in the 4D-Var framework. This cost-function is usually solved by a sequence of quadratic minimizations using linearized operators, an algorithm known in meteorology as incremental 4D-Var (Courtier et al., 1994). Here we are only interested in the final value of the cost function (at the minimum), but of course preconditioning in minimizing Eq. (1) is a crucial problem in practice (e.g. Lorenc, 1997; Gauthier et al., 1999). A Bayesian derivation of Eq. (1) is provided by Lorenc (1986).

Introducing the innovation vector

$$\mathbf{d} = \mathbf{y}_o - \mathcal{H}(\mathbf{x}_b) \quad (2)$$

the linearized cost-function (still written J) is expressed as:

$$\begin{aligned} \delta_x &= \mathbf{x} - \mathbf{x}_b \\ J(\delta_x) &= J_b(\delta_x) + J_o(\delta_x) \\ &= \frac{1}{2} \{ \delta_x^T \mathbf{B}^{-1} \delta_x + [\mathbf{d} - \mathbf{H}\delta_x]^T \mathbf{R}^{-1} [\mathbf{d} - \mathbf{H}\delta_x] \} \end{aligned} \quad (3)$$

where \mathbf{H} is the linearized observation operator and where every term is now quadratic.

A common derivation of the Bennett-Talagrand criterion casts the problem (3) into a single term using a generalized observation operator (Talagrand, 1999; Desroziers and Ivanov, 2001). This is not necessary for deriving the statistical moments of $\tilde{J} = J(\mathbf{x}_a)$. An alternative way is to write down the solution of Eq. (3) in the standard form involving the data assimilation operator (or gain matrix) \mathbf{K} :

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}\mathbf{d} \quad (4)$$

\mathbf{K} is a representation of performing variational assimilation for the analysis increments. The background term at the minimum \tilde{J}_b is therefore:

$$\tilde{J}_b = \frac{1}{2} \mathbf{d}^T \mathbf{K}^T \mathbf{B}^{-1} \mathbf{K} \mathbf{d} \quad (5)$$

and similarly, the observation term at the minimum \tilde{J}_o is:

$$\tilde{J}_o = \frac{1}{2} \mathbf{d}^T (\mathbf{I} - \mathbf{H}\mathbf{K})^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{H}\mathbf{K}) \mathbf{d} \quad (6)$$

From Eq. (3), it is also possible to express the observation term before the minimization \hat{J}_o as

$$\hat{J}_o = \frac{1}{2} \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d} \quad (7)$$

and $\hat{J}_b = 0$. This explicitly shows that these terms are (halves of) quadratic forms in the innovation vector \mathbf{d} .

In general, observation errors are taken to be diagonal or block-diagonal, with each block corresponding to a given instrument, and observation errors are taken to be uncorrelated

between instruments. Following Chapnik et al. (2004), this is written as a projection operator Π_i^o operating on the observation vector \mathbf{y} and giving a subset $\mathbf{y}_i = \Pi_i^o \mathbf{y}$ of the observations, with $i = 1, \dots, P$ and where P is the number of such subsets. Up to a reordering, the matrix \mathbf{R} can be written in the block-diagonal form:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_P \end{pmatrix}$$

where $\mathbf{R}_i = \Pi_i^o \mathbf{R} \Pi_i^{oT}$ are observation error covariance matrices for the i th-subset of observations. The inverse covariance matrix \mathbf{R}^{-1} is also block-diagonal, with elements \mathbf{R}_i^{-1} , and the observation cost-function can be split into P contributions J_o^i (either at the minimum or before the minimization):

$$\tilde{J}_o^i = \frac{1}{2} \mathbf{d}^T (\mathbf{I} - \mathbf{H}\mathbf{K})^T \Pi_i^{oT} \mathbf{R}_i^{-1} \Pi_i^o (\mathbf{I} - \mathbf{H}\mathbf{K}) \mathbf{d} \quad (8)$$

$$\hat{J}_o^i = \frac{1}{2} \mathbf{d}^T \Pi_i^{oT} \mathbf{R}_i^{-1} \Pi_i^o \mathbf{d} \quad (9)$$

which are quadratic forms in the innovation vector.

In general the background error covariance matrix is not block-diagonal as the different variables are correlated. This was neglected in DI01, but it can be taken into account by introducing the parameter transform \mathbf{K}_p that changes variables into uncorrelated ones (Derber and Bouttier, 1999; Bannister, 2008). Different spatial autocovariance blocks \mathbf{B}_i are specified for each unbalanced variable. This can be written as:

$$\begin{aligned} \mathbf{B} &= \mathbf{K}_p \begin{pmatrix} \mathbf{B}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}_S \end{pmatrix} \mathbf{K}_p^T \\ &= \mathbf{K}_p \sum_{i=1}^{i=S} \Pi_i^{bT} \mathbf{B}_i \Pi_i^b \mathbf{K}_p^T \end{aligned}$$

where usually there are $S=3-5$ control variables (usually vorticity or streamfunction, unbalanced divergence or velocity potential, and some form of unbalanced temperature and humidity, see Bannister (2008) for a review and comparison between various implementations). \mathbf{K}_p is generally formulated as a lower-triangular non-singular matrix with left-inverse \mathbf{K}_p^{-1} . The projection matrix Π_i^b is used to extract the i th-component of the background vector in unbalanced space and can be written explicitly with blocks being either the zero matrix $\mathbf{0}$ or the identity matrix \mathbf{I} in i th position:

$$\Pi_i^b = (\mathbf{0} \dots \mathbf{0} \mathbf{I} \mathbf{0} \dots \mathbf{0}).$$

Using this formulation allows us to split the background term into S contributions J_b^i that are, again, quadratic forms in \mathbf{d} :

$$\begin{aligned}\tilde{J}_b &= \sum_{i=1}^S \tilde{J}_b^i \\ \tilde{J}_b^i &= \frac{1}{2} \mathbf{d}^T \mathbf{K}^T \mathbf{K}_p^{-T} \Pi_i^{bT} \mathbf{B}_i^{-1} \Pi_i^b \mathbf{K}_p^{-1} \mathbf{K} \mathbf{d}\end{aligned}\quad (10)$$

There may be additional penalty terms in the cost function, related for instance to the coupling information in a limited area model (Guidard and Fischer, 2008) or to the weak constraint digital filter (Gauthier and Thépaut, 2001). These terms will not be considered explicitly here, but they can be put into similar quadratic forms. The digital filter expresses a constraint on the increments which are directly related to the innovation vector. The coupling information requires an extended definition of the information vector, and this issue may be addressed in future work.

2.2 Moments

The innovation vector is a random variable, so are the background and observation terms at the minimum, and it is possible to explicitly derive their expectation, but also their variances and covariances in the Gaussian case. Note that this Gaussian assumption can be relaxed for some classes of distributions; see e.g. Mathai et al. (1995) for the case of elliptically contoured distributions and Genton et al. (2001) for skew-normal ones.

2.2.1 Expectations

Without any loss of generality in this section (other than existence of second order moments), we will denote the first two moments of the innovation vector by:

$$\begin{aligned}\boldsymbol{\mu} &= E(\mathbf{d}) \\ \boldsymbol{\Sigma} &= \text{Cov}(\mathbf{d})\end{aligned}$$

where $E(\cdot)$ is the expectation and $\text{Cov}(\cdot)$ the covariance. Here, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the actual (i.e. measured) mean and covariance of the innovation vector that may differ from their assumed values given in Eqs. (14) and (15). Then, it can be seen by applying the trace operator (M92, Corollary 3.2b.1) that the expectation of any quadratic form in \mathbf{d} is:

$$E(\mathbf{d}^T \mathbf{A} \mathbf{d}) = \text{Tr}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}\quad (11)$$

where $\text{Tr}(\cdot)$ is the trace (which is a linear operator). This result is robust as it holds for any distribution of \mathbf{d} with finite first-two moments (Ménard et al., 2000). Equation (11) can be readily applied to the previously derived quadratic forms. In particular the expectation of background and observation

terms follows:

$$E(\tilde{J}_b) = \frac{1}{2} \text{Tr}(\mathbf{K}^T \mathbf{B}^{-1} \mathbf{K} \boldsymbol{\Sigma}) + \frac{1}{2} \boldsymbol{\mu}^T \mathbf{K}^T \mathbf{B}^{-1} \mathbf{K} \boldsymbol{\mu}\quad (12)$$

$$\begin{aligned}E(\tilde{J}_o) &= \frac{1}{2} \text{Tr}[(\mathbf{I} - \mathbf{H} \mathbf{K})^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{H} \mathbf{K}) \boldsymbol{\Sigma}] \\ &\quad + \frac{1}{2} \boldsymbol{\mu}^T (\mathbf{I} - \mathbf{H} \mathbf{K})^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{H} \mathbf{K}) \boldsymbol{\mu}\end{aligned}\quad (13)$$

No hypothesis has been made on the distribution of the innovation yet. These equations are valid for any data assimilation scheme with possibly incorrect error covariance matrices (and hence \mathbf{K}).

Now, in a well specified system, we would have unbiased innovation and well-prescribed statistics:

$$\boldsymbol{\mu} = \mathbf{0}\quad (14)$$

$$\boldsymbol{\Sigma} = \mathbf{D} \equiv \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}\quad (15)$$

The gain matrix can also be written as:

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

which allows us to write the following expressions:

$$\mathbf{B}^{-1} \mathbf{K} \mathbf{D} = \mathbf{H}^T$$

$$\mathbf{R}^{-1} (\mathbf{I} - \mathbf{H} \mathbf{K}) \mathbf{D} = \mathbf{I}$$

Thus, when $\boldsymbol{\Sigma} = \mathbf{D}$, the expectation of background and observation terms take the form

$$E(\tilde{J}_b) = \frac{1}{2} \text{Tr}[\mathbf{K}^T \mathbf{H}^T] = \frac{1}{2} \text{Tr}(\mathbf{H} \mathbf{K})\quad (16)$$

$$E(\tilde{J}_o) = \frac{1}{2} \text{Tr}[(\mathbf{I} - \mathbf{H} \mathbf{K})^T] = \frac{1}{2} \text{Tr}(\mathbf{I} - \mathbf{H} \mathbf{K})\quad (17)$$

as previously derived by Bennett et al. (1993), Talagrand (1999), and Desroziers and Ivanov (2001).

New results are obtained by applying a similar calculation with sub-terms of the background cost function at the minimum:

$$\begin{aligned}E(\tilde{J}_b^i) &= \frac{1}{2} \text{Tr}[\mathbf{K}^T \mathbf{K}_p^{-T} \Pi_i^{bT} \mathbf{B}_i^{-1} \Pi_i^b \mathbf{K}_p^{-1} \mathbf{B} \mathbf{H}^T] \\ &= \frac{1}{2} \text{Tr}\left[\mathbf{H}^T \mathbf{K}^T \mathbf{K}_p^{-T} \Pi_i^{bT} \mathbf{B}_i^{-1} \Pi_i^b \sum_{j=1}^{i=S} \Pi_j^{bT} \mathbf{B}_j \Pi_j^b \mathbf{K}_p^T\right] \\ &= \frac{1}{2} \text{Tr}[\mathbf{H}^T \mathbf{K}^T \mathbf{K}_p^{-T} \Pi_i^{bT} \Pi_i^b \mathbf{K}_p^T] \\ &= \frac{1}{2} \text{Tr}[\mathbf{K}_p \Pi_i^{bT} \Pi_i^b \mathbf{K}_p^{-1} \mathbf{K} \mathbf{H}].\end{aligned}\quad (18)$$

The expectation of any sub-term of the observation cost function at the minimum is derived in a similar way:

$$E(\tilde{J}_o^i) = \frac{1}{2} \text{Tr}[\Pi_o^i{}^T \Pi_o^i (\mathbf{I} - \mathbf{H} \mathbf{K})].\quad (19)$$

Analogous expressions can be obtained for the sub-term of the observation cost function before the minimization (Eqs. 7 and 9).

2.2.2 Variances

In general, the variance of the quadratic forms will depend on the fourth-order moment of \mathbf{d} . If the innovation vector has a Gaussian distribution, then its fourth-order moment can be expressed as a function of the first two ones, and the variance of the quadratic forms can be expressed following M92 (theorem 3.2b.2):

$$\text{Var}(\mathbf{d}^T \mathbf{A} \mathbf{d}) = 2\text{Tr}[(\mathbf{A}\Sigma)^2] + 4\boldsymbol{\mu}^T \mathbf{A} \Sigma \mathbf{A} \boldsymbol{\mu}. \quad (20)$$

This formula can be readily applied to the previous quadratic forms (without assuming perfect statistics). Now, in the particular case of a system with well-specified statistics verifying Eqs. (14)–(15) we have:

$$\text{Var}(\tilde{J}_b) = \frac{1}{2} \text{Tr}[(\mathbf{H}\mathbf{K})^2] \quad (21)$$

$$\text{Var}(\tilde{J}_o) = \frac{1}{2} \text{Tr}[(\mathbf{I} - \mathbf{H}\mathbf{K})^2]. \quad (22)$$

The results do not seem to appear in the standard literature on data assimilation for the atmosphere, yet analogous formulas appear in Bennett et al. (2000) within the formalism of the representer method. They can be extended to sub-parts of the cost function (not shown).

Before deriving cross-covariances in the next section, one can point out that \tilde{J}_b and \tilde{J}_o should be positively correlated. Indeed, using the fact that $2\tilde{J}$ has a χ_p^2 distribution, we can deduce that:

$$\begin{aligned} \text{Var}(\tilde{J}) &= p/2 \\ &= \text{Var}\tilde{J}_b + \text{Var}\tilde{J}_o + 2\text{Cov}(\tilde{J}_b, \tilde{J}_o). \end{aligned}$$

The matrices $\mathbf{H}\mathbf{K}$ and $\mathbf{I} - \mathbf{H}\mathbf{K}$ have eigenvalues between 0 and 1 (?), therefore:

$$\begin{aligned} \text{Tr}[(\mathbf{H}\mathbf{K})^2] + \text{Tr}[(\mathbf{I} - \mathbf{H}\mathbf{K})^2] \\ \leq \text{Tr}(\mathbf{H}\mathbf{K}) + \text{Tr}(\mathbf{I} - \mathbf{H}\mathbf{K}) = \text{Tr}(\mathbf{I}) = p. \end{aligned}$$

Thus it is necessary that

$$\text{Cov}(\tilde{J}_b, \tilde{J}_o) \geq 0. \quad (23)$$

In other words, in a system with well-specified statistics, \tilde{J}_b and \tilde{J}_o are positively correlated. This may sound like a counter-intuitive result, as one could expect the random fluctuations of J to be “split” (with anticorrelations) between \tilde{J}_b and \tilde{J}_o , but this is not the case. Rather, the random fluctuations go in the same direction.

2.2.3 Covariances

The covariance between (sub)terms of the cost function can be calculated as follows (M92, theorem 3.2d.4):

$$\begin{aligned} \text{Cov}(\mathbf{d}^T \mathbf{A}_1 \mathbf{d}, \mathbf{d}^T \mathbf{A}_2 \mathbf{d}) &= 2\text{Tr}(\mathbf{A}_1 \Sigma \mathbf{A}_2 \Sigma) \\ &\quad + 4\boldsymbol{\mu}^T \mathbf{A}_1 \Sigma \mathbf{A}_2 \boldsymbol{\mu} \end{aligned} \quad (24)$$

In particular, in a well-specified system:

$$\text{Cov}(\tilde{J}_b, \tilde{J}_o) = \frac{1}{2} \text{Tr}[\mathbf{H}\mathbf{K}(\mathbf{I} - \mathbf{H}\mathbf{K})] \quad (25)$$

The magnitude of the correlation depends on how $\text{Tr}(\mathbf{H}\mathbf{K})$ and $\text{Tr}[(\mathbf{H}\mathbf{K})^2]$ differ. This highlights that it could be useful to estimate $\text{Tr}[(\mathbf{H}\mathbf{K})^2]$ in operational schemes. Also, further expressions can be derived for the cross-covariances between sub-parts of the cost function (not shown).

2.3 Distribution

Beyond the statistical moments, it may be of some interest to get information on the distribution of the sub-parts of the cost function at the minimum. The random variable $2\tilde{J}$ has a variance that is double of its expectation, as expected for a χ_p^2 -distributed variable. In view of Eqs. (16)–(21) and (17)–(22), this is generally not the case for $2\tilde{J}_o$, $2\tilde{J}_b$ (nor their sub-parts). Thus they do not follow a χ^2 distribution – see also the necessary and sufficient conditions in the chapter 5 of M92. However, they are linear combinations of χ^2 distributions in the Gaussian case (M92).

Albeit the previous expressions may be thought to be restricted to variational assimilation, they would also apply to an ensemble variant of the Kalman filter provided that the associated cost function can be computed (Ménard and Chang, 2000; Talagrand, 2010).

3 Practical estimation and potential use

This section introduces the randomized trace algorithm to compute the previous statistical expressions. They could be used for hypothesis testing. Rather, we focus on their use when tuning global scaling factors for matrices \mathbf{B} and \mathbf{R} .

3.1 Application of the randomized trace estimator

The previous expressions involve the computation of the trace of large matrices that are unavailable in large-size data assimilation schemes used in operations in meteorology. Following DI01, it is however possible to estimate them with a randomized trace estimator:

$$\text{Tr}(\mathbf{A}) \simeq \frac{1}{M} \sum_{i=1}^M \boldsymbol{\eta}_i^T \mathbf{A} \boldsymbol{\eta}_i \quad (26)$$

where $\boldsymbol{\eta}_i$ are independent random vectors whose components follow a standard $\mathcal{N}(0, 1)$ normal distribution (Girard, 1991) or Rademacher (± 1 with probability 1/2) distribution (Hutchinson, 1989). The Hutchinson estimator has lower variance than Girard’s but requires more samples to achieve a given relative precision (Avron and Toledo, 2011). In addition, using the Gaussian distribution is more natural in data assimilation and allows $\text{Tr}(\mathbf{H}\mathbf{K})$ to be computed as a by-product of an ensemble of variational assimilations

(Desroziers et al., 2009). It is shown next that randomized estimation of traces in new formulas (21,22, and 25) is possible through perturbed minimizations (as in DI01), yet at the cost of an additional minimization. Indeed, \mathbf{K} (the assimilation algorithm) has to be applied three times.

Instead of the direct application of Eq. (26), DI01 take the approach of further introducing the square root forms of the error covariance matrices $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$ and $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{T/2}$. Then, using the fact that the trace is invariant under cyclic permutations (not general ones), the randomized estimate (26) required in Eqs. (21), (22), and (25) follows:

$$\begin{aligned} \text{Tr}[(\mathbf{HK})^2] &= \text{Tr}(\mathbf{HKHKR}^{1/2}\mathbf{R}^{T/2}\mathbf{R}^{-1}) \\ &= \text{Tr}(\mathbf{R}^{T/2}\mathbf{R}^{-1}\mathbf{HKHKR}^{1/2}) \\ &\simeq \frac{1}{M} \sum_{i=1}^M (\mathbf{R}^{1/2}\boldsymbol{\eta}_i)^T \mathbf{R}^{-1}\mathbf{HKHK}(\mathbf{R}^{1/2}\boldsymbol{\eta}_i) \\ &= \frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{y}_i^{\circ T} \mathbf{R}^{-1}\mathbf{HKHK}\boldsymbol{\delta}\mathbf{y}_i^{\circ} \\ &= \frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{y}_i^{\circ T} \mathbf{R}^{-1}\mathbf{HKH}(\boldsymbol{\delta}\mathbf{x}_a^* - \boldsymbol{\delta}\mathbf{x}_{a|y_0}) \\ &= \frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{y}_i^{\circ T} \mathbf{R}^{-1}\mathbf{H}(\boldsymbol{\delta}\mathbf{x}_a^{**} - \boldsymbol{\delta}\mathbf{x}_{a|y_0}) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\delta}\mathbf{x}_a^* &= \boldsymbol{\delta}\mathbf{x}_{a|y_0+\boldsymbol{\delta}\mathbf{y}_i^{\circ}} \\ \boldsymbol{\delta}\mathbf{x}_a^{**} &= \boldsymbol{\delta}\mathbf{x}_{a|y_0+\mathbf{H}(\boldsymbol{\delta}\mathbf{x}_a^*-\boldsymbol{\delta}\mathbf{x}_{a|y_0})} \end{aligned} \quad (27)$$

and $\boldsymbol{\delta}\mathbf{y}_i^{\circ} = \mathbf{R}^{1/2}\boldsymbol{\eta}_i$ are sets of perturbations of observations (accordingly to the error covariance matrix \mathbf{R}). Thus, at least three $(2M+1)$ minimizations are necessary for the computation of this trace: first with original observations to give the unperturbed analysis increment $\boldsymbol{\delta}\mathbf{x}_{a|y_0}$, then with perturbed observations to give the perturbed analysis increment $\boldsymbol{\delta}\mathbf{x}_a^*$, and finally the difference between those analysis increments is brought into observation space and used as a perturbation of observation in a third analysis giving the analysis increment $\boldsymbol{\delta}\mathbf{x}_a^{**}$. The trace computation then follows, provided that \mathbf{R}^{-1} is available as an operator, which is easy if observations errors are taken uncorrelated as it is usually the case.

DI01 have provided an independent computation of $\text{Tr}(\mathbf{KH}) = \text{Tr}(\mathbf{HK})$. It is also possible to derive a similar computation of $\text{Tr}[(\mathbf{HK})^2] = \text{Tr}[(\mathbf{KH})^2]$, as:

$$\begin{aligned} \text{Tr}[(\mathbf{KH})^2] &= \text{Tr}(\mathbf{KHKHB}^{1/2}\mathbf{B}^{T/2}\mathbf{B}^{-1}) \\ &= \text{Tr}(\mathbf{B}^{T/2}\mathbf{B}^{-1}\mathbf{KHKHB}^{1/2}) \\ &\simeq \frac{1}{M} \sum_{i=1}^M (\mathbf{B}^{1/2}\boldsymbol{\epsilon}_i)^T \mathbf{B}^{-1}\mathbf{KHKH}(\mathbf{B}^{1/2}\boldsymbol{\epsilon}_i) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{x}_i^{\text{b}T} \mathbf{B}^{-1}\mathbf{KHKH}\boldsymbol{\delta}\mathbf{x}_i^{\text{b}} \\ &= -\frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{x}_i^{\text{b}T} \mathbf{B}^{-1}\mathbf{KH}(\boldsymbol{\delta}\mathbf{x}_a^{\circ} - \boldsymbol{\delta}\mathbf{x}_{a|x^{\text{b}}}) \\ &= -\frac{1}{M} \sum_{i=1}^M \boldsymbol{\delta}\mathbf{x}_i^{\text{b}T} \mathbf{B}^{-1}(\boldsymbol{\delta}\mathbf{x}_a^{\circ\circ} - \boldsymbol{\delta}\mathbf{x}_{a|x^{\text{b}}}) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\delta}\mathbf{x}_a^{\circ} &= \boldsymbol{\delta}\mathbf{x}_{a|x^{\text{b}}+\boldsymbol{\delta}\mathbf{x}_i^{\text{b}}} \\ \boldsymbol{\delta}\mathbf{x}_a^{\circ\circ} &= \boldsymbol{\delta}\mathbf{x}_{a|x^{\text{b}}+\boldsymbol{\delta}\mathbf{x}_a^{\circ}-\boldsymbol{\delta}\mathbf{x}_{a|x^{\text{b}}}} \end{aligned} \quad (28)$$

and $\boldsymbol{\delta}\mathbf{x}_i^{\text{b}} = \mathbf{B}^{1/2}\boldsymbol{\epsilon}_i$ are sets of background perturbations. Thus, at least three minimizations are necessary for the computation of this trace: first with original background; second with a perturbed background, third with a background perturbed by differences of analysis increments. This last formula may be simplified when working with $\mathbf{B}^{1/2}$ -preconditioning (e.g. Bannister, 2008 as it can be expressed in terms of the control variable. Also, those computations only involve analysis increments where observations or background have been perturbed, and they can still be computed when using a weakly non-linear algorithm.

Finally, another way to compute directly the expectations and covariances of the sub-parts of the cost function is the “simulated optimal innovation technique” introduced by Chapnik et al. (2006). In this approach, pseudo-observations and pseudo-backgrounds are generated with errors that are exactly consistent with the prescribed error statistics. Then, assimilation of these synthetic observations will result into sub-parts of the cost function that exactly follow the previous expressions. This is a Monte-Carlo approach to compute the expectations and covariances in the specific case where assimilation statistics are correctly specified.

3.2 A simplified framework

If one wants to compare the theoretical statistics to the observed ones, the obvious problem lies in the fact that there is only one realization of the observations (and thus of the innovation vector) in the real atmosphere. The expectation of the cost-function can be estimated from a single realization (Chapnik et al., 2004), yet for some observation types the use of space or time averages to build samples large enough is required (Sadiki and Fischer, 2005). Obviously, computing the variance of the cost-function requires many samples, thus the use of space or time averages. This kind of ergodicity assumption raises questions in itself, such that we will start investigating the use of these formulas first in a simplified framework where independent Monte-Carlo realizations of the innovation vector can be easily generated. Such an approach was also used in Muccino et al. (2004) with the same purpose.

For that purpose, a simplified one-dimensional variational assimilation framework has been set up in a similar way as in DI01. The goal is to analyze a single variable (say temperature) on a circular domain (say on an earth meridian), minimizing Eq. (3). A conjugate gradient algorithm with $\mathbf{B}^{1/2}$ -preconditioning is used for the minimization. It involves a maximum of 100 iterations, which is sufficient for convergence unless correlated observation errors are specified (this is verified a posteriori). Background errors have homogeneous standard deviation $\sigma_b = 1$ K. The correlation function c is chosen to belong to the Matérn class (Rasmussen and Williams, 2006) with “smoothness” parameter $\nu = 1$, corresponding to:

$$c(x) = \left(1 + \frac{\sqrt{3}x}{\ell}\right) \exp\left(-\frac{\sqrt{3}x}{\ell}\right)$$

where the length-scale is chosen to be $\ell = 250$ km; see Michel (2013) for details. The observation operator is a periodic linear interpolation. Observations are taken to be equally spaced, which allows easy inclusion of spatial correlations if needed, using the same model as \mathbf{B} but in observation space. In order to roughly simulate the characteristics of operational data assimilation, the number of variables is taken as $n = 1000$ and there are $p = 500$ observations with precision $\sigma_o = 1$ K. The fields are illustrated in Fig. 1. In this particular case, the value of the cost-function at the minimum is $\tilde{J} = 245.8 \approx p/2$ as the specified error statistics are fully consistent with the ones of the innovation vector.

3.3 Validation of the randomized trace estimation

The distribution of the observation and background terms at the minimum have been obtained from running 10^4 independent realizations of the assimilation scheme. The distribution of the obtained values is shown in Fig. 2. The total cost function has a mean of ≈ 250.11 which is consistent with the theoretical expectation $p/2 = 250$ (within the bounds of a standard confidence interval with 10^4 realizations). The estimated variances of the observation and background terms are lower than their mean, which is consistent with the discussion in Sect. 2.

The trace estimation algorithm relies on randomization, thus its results depend on the sample size. The algorithm has been applied to compute $\text{Tr}(\mathbf{HK})$, $\text{Tr}[(\mathbf{HK})^2]$, and also independently $\text{Tr}(\mathbf{KH})$ and $\text{Tr}[(\mathbf{KH})^2]$ for various ensemble (sampling) sizes M between 1 and 100. The procedure has been applied 200 times in order to evaluate the statistical properties of the randomized trace estimator. Results are presented in Fig. 3. Almost identical results are obtained among ensemble sizes. Averaging over the ensemble sizes and the 200 realizations gives the estimations $\text{Tr}(\mathbf{HK}) \approx 80.15$, $\text{Tr}(\mathbf{KH}) \approx 80.19$, $\text{Tr}[(\mathbf{HK})^2] \approx 56.7$, $\text{Tr}[(\mathbf{KH})^2] \approx 56.86$ which are obviously consistent whenever Eq. (26) or Eq. (27) is used. The comparison with

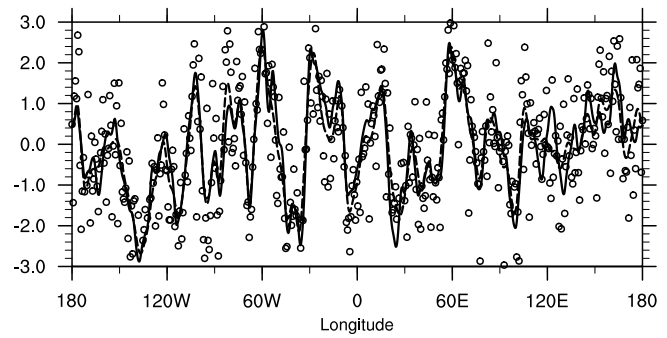


Fig. 1. True signal x_t (solid black line), observations y_o (circles) and the analysis x_a (dashed black line) simulated with the prescribed statistics for background and observation errors (see text). The background x_b is zero.

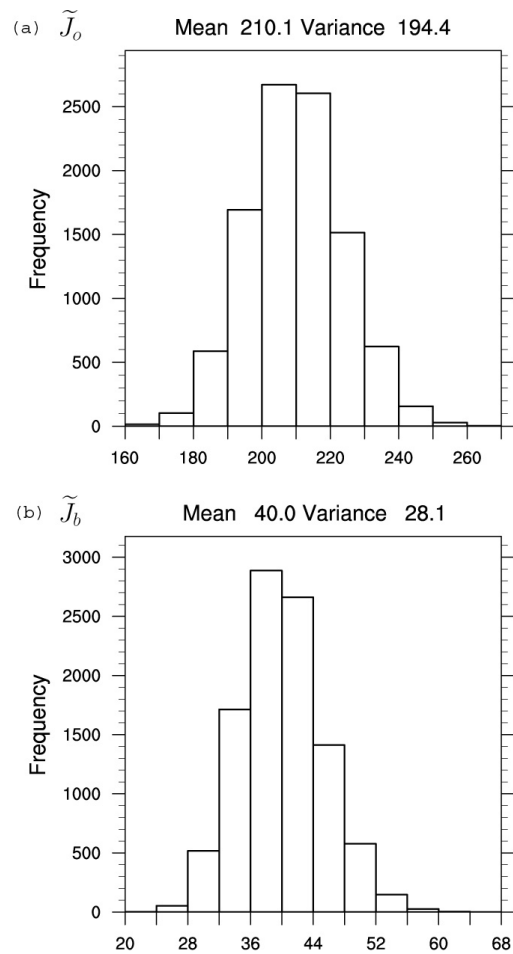


Fig. 2. Histograms obtained with 10^4 realizations of the assimilation scheme shown in Fig. 1 for the values of (a) the background cost function and (b) the observation cost function at the minimum.

the independent realizations of the assimilation scheme (the simulated optimal innovation method, Fig. 2) is shown in Table 1.

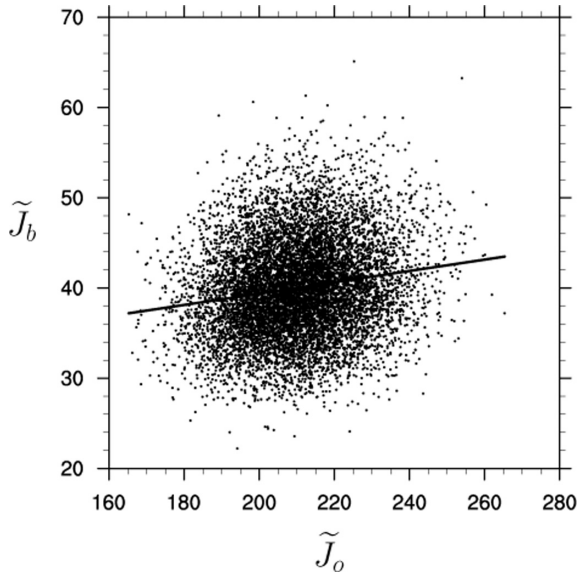


Fig. 3. Scatterplot of the values of the background cost function and the observation cost function at the minimum.

Table 1. Comparison of the simulated optimal innovation method with the randomized trace estimation, following Eqs. (16), (17), (21), and (22).

$E(\tilde{J}_b)$	40.01	$\frac{1}{2}\text{Tr}(\mathbf{HK})$	40.09
$\text{Var}(\tilde{J}_b)$	28.14	$\frac{1}{2}\text{Tr}[(\mathbf{HK})^2]$	28.41
$E(\tilde{J}_o)$	210.11	$\frac{1}{2}\text{Tr}(\mathbf{I} - \mathbf{HK})$	209.91
$\text{Var}(\tilde{J}_o)$	194.38	$\frac{1}{2}\text{Tr}[(\mathbf{I} - \mathbf{HK})^2]$	198.24

Differences between theoretical and estimated values arise from sample size. Moving to cross-correlation between the values of \tilde{J}_o and \tilde{J}_b , the simulated optimal innovation method gives an estimate

$$\text{Corr}(\tilde{J}_o, \tilde{J}_b) \simeq 0.164 \quad (29)$$

which is also in rather good agreement with the estimation from the trace algorithm:

$$\frac{\text{Tr}[\mathbf{HK}(\mathbf{I} - \mathbf{HK})]}{\sqrt{\text{Tr}[(\mathbf{HK})^2]\text{Tr}[(\mathbf{I} - \mathbf{HK})^2]}} \simeq 0.156 \quad (30)$$

as also shown in Fig. 3; \tilde{J}_o and \tilde{J}_b are positively correlated.

Figure 4 shows how the trace estimator becomes more and more precise as the ensemble size goes larger. More precisely, the variance of the estimator (26) is (Chapnik et al., 2006):

$$\text{Var}\left(\frac{1}{M} \sum_{i=1}^M \eta_i^T \mathbf{A} \eta_i\right) = \frac{1}{M} \text{Tr}\left[\left(\frac{\mathbf{A} + \mathbf{A}^T}{2}\right)^2\right].$$

Thus, it is possible to estimate the precision of the randomized trace computation as a function of ensemble size

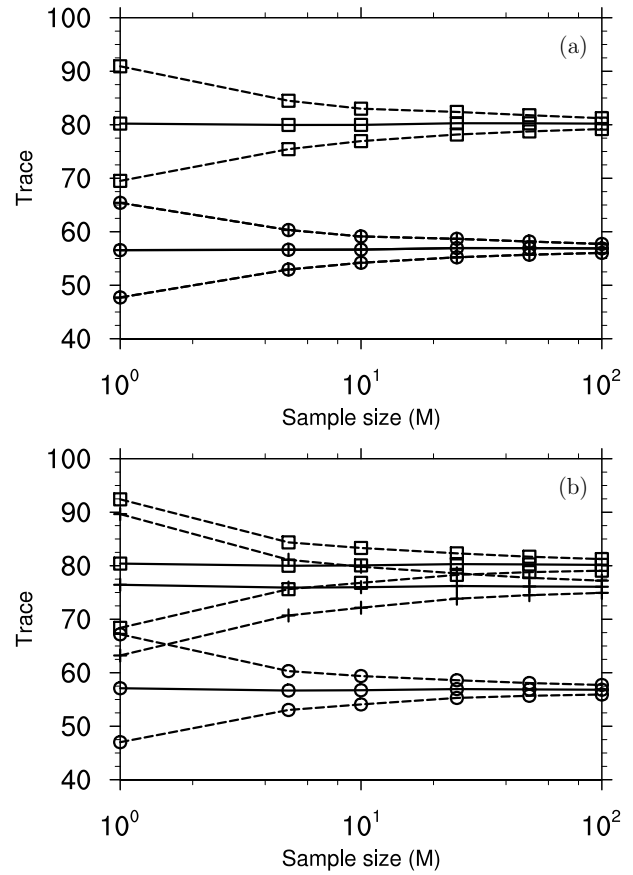


Fig. 4. Randomized trace estimation of $\text{Tr}(\mathbf{HK})$ (a, squares), $\text{Tr}[(\mathbf{HK})^2]$ (a, circles), $\text{Tr}[(\mathbf{HK})^T \mathbf{HK}]$ (a, plus signs) and $\text{Tr}(\mathbf{KH})$ (b, squares), $\text{Tr}[(\mathbf{KH})^2]$ (b, circles), $\text{Tr}[(\mathbf{KH})^T \mathbf{KH}]$ (b, plus signs), as a function of the ensemble size used for the randomization. Solid line: expectation; dashed lines: expectation plus or minus one standard deviation, as determined from a Monte-Carlo approach with 200 independent realizations.

M and as a function of traces $\text{Tr}[(\mathbf{HK})^2]$, $\text{Tr}[(\mathbf{HK})^T \mathbf{HK}]$, $\text{Tr}[(\mathbf{KH})^2]$ and $\text{Tr}[(\mathbf{KH})^T \mathbf{KH}]$ which themselves can be computed with the randomized approach proposed in Sect. 3.1. This also implies that while Eqs. (25) and (26) can be applied to estimate $\text{Tr}(\mathbf{HK}) = \text{Tr}(\mathbf{KH})$, both estimators have different precisions, as apparent in Fig. 4.

3.4 Variance tuning

DI01 proposed to adjust multiplicative scaling factors s^b and s^o for the covariance matrices \mathbf{B} and \mathbf{R} , assuming that at least approximately

$$\mathbf{B}_t = s^b \mathbf{B} \quad (31)$$

$$\mathbf{R}_t = s^o \mathbf{R} \quad (32)$$

where \mathbf{B}_t and \mathbf{R}_t are the true covariance matrices. Then DI01 proposed to estimate the scaling factors as the ratio of

the realization of the cost function at the minimum to their expectations:

$$s^b = \frac{2\tilde{J}_b}{\text{Tr}(\mathbf{HK})} \quad (33)$$

$$s^o = \frac{2\tilde{J}_o}{\text{Tr}(\mathbf{I} - \mathbf{HK})} \quad (34)$$

These equations can be seen as a fixed point relation that is solved iteratively.

Alternatively, if the conditions (31)–(32) are satisfied, then it is also possible to derive an alternative formulation for those coefficients. Indeed, the innovation vector \mathbf{d} has covariance

$$\begin{aligned} \Sigma &= \mathbf{HB}_t\mathbf{H}^T + \mathbf{R}_t = s^b\mathbf{HBH}^T + s^o\mathbf{R} \\ &= (s^b - s^o)\mathbf{HBH}^T + s^o\mathbf{D} \end{aligned} \quad (35)$$

$$= s^b\mathbf{D} + (s^o - s^b)\mathbf{R}. \quad (36)$$

Using Eq. (35) in Eq. (12) with $\boldsymbol{\mu} = 0$ gives:

$$\begin{aligned} 2\mathbf{E}(\tilde{J}_b) &= \text{Tr}(\mathbf{K}^T\mathbf{B}^{-1}\mathbf{K}\Sigma) \\ &= s^b\text{Tr}(\mathbf{K}^T\mathbf{B}^{-1}\mathbf{K}\mathbf{D}) + (s^o - s^b)\text{Tr}[\mathbf{K}^T\mathbf{H}^T\mathbf{D}^{-1}\mathbf{R}] \\ &= s^b\text{Tr}(\mathbf{HK}) + (s^o - s^b)\text{Tr}[\mathbf{HK}(\mathbf{I} - \mathbf{HK})] \end{aligned} \quad (37)$$

Using Eq. (36) in Eq. (13) with $\boldsymbol{\mu} = 0$ gives

$$2\mathbf{E}(\tilde{J}_o) = s^o\text{Tr}(\mathbf{I} - \mathbf{HK}) + (s^b - s^o)\text{Tr}[\mathbf{HK}(\mathbf{I} - \mathbf{HK})] \quad (38)$$

which was already been obtained by Chapnik et al. (2006) when discussing the first iteration of the algorithm designed by DI01. In view of Eqs. (21)–(22), the random fluctuations of \tilde{J}_b and \tilde{J}_o are small and the expectation can usually be replaced by a single realization. This turns Eqs. (37)–(38) into a *linear system* of two unknowns s^o, s^b which can be written as:

$$\alpha s^b + (\alpha - \beta)(s^o - s^b) = 2\tilde{J}_b$$

$$(p - \alpha)s^o + (\alpha - \beta)(s^b - s^o) = 2\tilde{J}_o$$

and where the notations $\alpha = \text{Tr}(\mathbf{HK})$, $\beta = \text{Tr}[(\mathbf{HK})^2]$ have been introduced for simplification. If the system is not singular, then the unique solution is:

$$s^b = 2 \frac{(\beta - \alpha)\tilde{J}_o + (p - 2\alpha + \beta)\tilde{J}_b}{\beta(p - 2\alpha + \beta) - (\beta - \alpha)^2} \quad (39)$$

$$s^o = 2 \frac{\beta\tilde{J}_o + (\beta - \alpha)\tilde{J}_b}{\beta(p - 2\alpha + \beta) - (\beta - \alpha)^2} \quad (40)$$

At the price of computing $\text{Tr}[(\mathbf{HK})^2]$, the final solution of DI01's algorithm can be directly obtained, without any further iteration in the case where specified and true error covariances differ by a scaling factor as indicated by Eqs. (31)–(32). This may however not be the case when observation operators are non-linear or when this condition is not met.

In our setting, $\text{Tr}[(\mathbf{HK})^2]$ is lower but comparable to $\text{Tr}(\mathbf{HK})$ in magnitude (see Sect. 3.3 for details). Both traces are smaller by about one order of magnitude with respect to the number of observations, which is consistent with recent estimates in ARPEGE (Desroziers et al., 2009, Fig. 1). DI01's formulation neglects the term $(s^o - s^b)\text{Tr}[\mathbf{HK}(\mathbf{I} - \mathbf{HK})]$ (but the correct estimate is recovered through the iterations). This term is indeed likely to be negligible in Eq. (38) but not so in Eq. (37). Thus, more iterations of the scheme may be needed to recover the appropriate value of s^b , rather than s^o , as already noticed by DI01 (Tables 3 and 4). In contrast, the correct estimate can be obtained by solving the linear system with the direct solution given by Eqs. (39)–(40). The iterative tuning of the coefficients following DI01's approach is shown in Fig. 5 for overestimated or underestimated σ_o and σ_b . They confirm that generally, a single iteration is necessary for tuning σ_o (panels a, c, e, and g). In contrast, around four iterations can be necessary for recovering a useful value of σ_b especially when $s^o - s^b$ is large (panels b and f), as the term $(s^o - s^b)\text{Tr}[\mathbf{HK}(\mathbf{I} - \mathbf{HK})]$ cannot be neglected. The formulation (39)–(40) recovers the correct factor within one iteration (shown only in panels a and b).

This suggests that it may be useful to use the formulation (39)–(40) when trying to diagnose a multiplicative factor for the background term. This happens for instance for the inflation of variances in an ensemble of data assimilations (Raynaud et al., 2012). Extension of this formulation to the simultaneous tuning of several parameters (corresponding to sub-parts of the background and the observation cost-functions) is possible but more complicated, as a linear system of order $S + P$ (the number of sub-terms) has to be solved. This will be the topic of further studies.

4 Application in the ARPEGE 4D-Var

This section describes the application of the variance tuning factors to the operational model of Météo-France. The global model ARPEGE is a spectral model with currently 70 vertical levels and features a stretched geometry that allows higher resolution over Europe (about 10 km in physical space). The assimilation scheme is based on the multi-incremental formulation (Courtier et al., 1994) with two outer-loops.

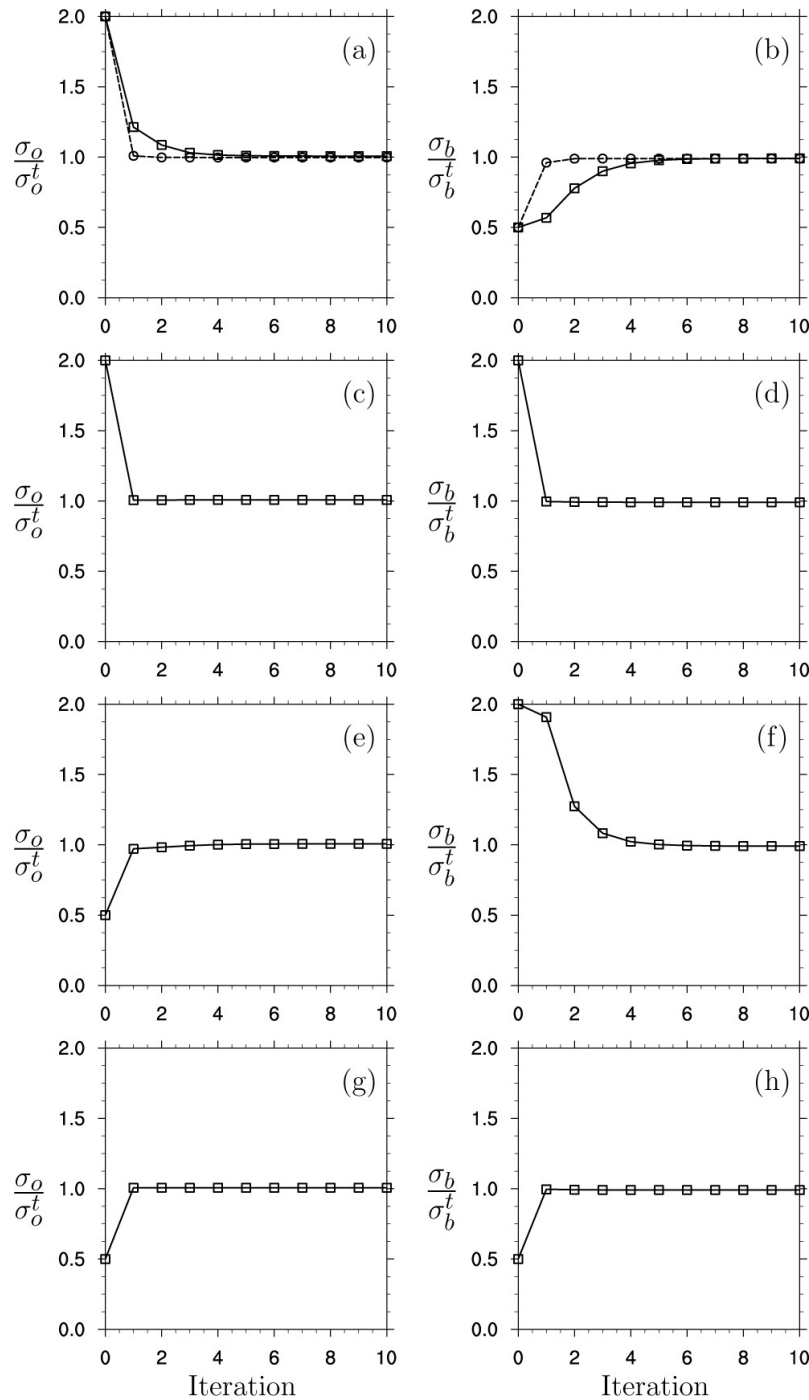


Fig. 5. Iterative tuning of multiplicative coefficients for the error covariance matrices. σ_o (left) and σ_b (right) are tuned simultaneously with **(a, b)** overestimated σ_o , underestimated σ_b ; **(c, d)** overestimated σ_o and σ_b ; **(e, f)** underestimated σ_o and overestimated σ_b ; **(g, h)** underestimated σ_o and $\sigma_b =$. Solid lines and square marks: iterative tuning following DI01's approach; dashed lines and circle marks: new approach (shown only in **a, b**). Traces are estimated with the randomization method using 100 independent realizations.

4.1 Estimation of tuning factors

Randomized estimates of the traces are performed in the same way as in the idealized 1-D model using the assimilation with perturbed observations. For the background term,

we estimate a single scaling factor with Eq. (39). Following the result of the previous section, it is clear that there is no need for such a refinement in the observation terms (e.g. inclusion of $(s^b - s^o)$ in Eq. 38). Observation scaling

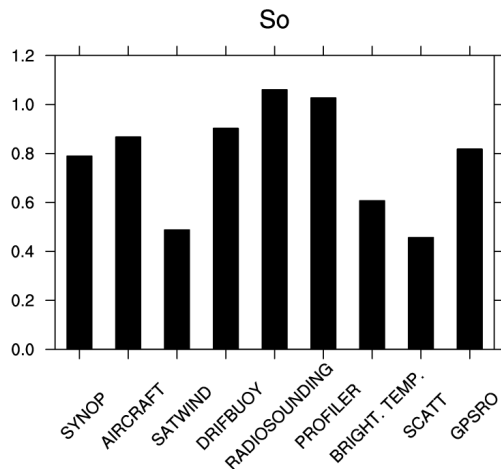


Fig. 6. Diagnosed σ_0 's-tuning factors per observation type (September 2013).

factors are therefore estimated in the very same way as in Desroziers et al. (2009). Randomized estimates of the traces have been averaged over a one week period and use a sample size $M = 6$, giving the following values:

$$\text{Tr}(\mathbf{HK}) \simeq 70\,000,$$

$$\text{Tr}[(\mathbf{HK})^2] \simeq 57\,000.$$

An application of the preceding expressions shows that background error standard deviations are overestimated by a factor 0.85. Thus, the specified \mathbf{B} matrix in ARPEGE is closer to its diagnosed value than in previous studies (Desroziers et al., 2009), but still $\mathbf{B} < \mathbf{B}_t$. The estimated values for the observation error standard deviations are shown in Fig. 6 for different observation types. Albeit previous studies have highlighted that observation errors are generally overestimated in data assimilation schemes (possibly because of approximate bias correction or neglect of spatial correlations in the errors), this was not the case anymore for some conventional observations (columns DRIFTBUOYS and RADIOSOUNDING have $S^o = \sqrt{s^o} > 1$). As explained by Desroziers et al. (2009), the obtained values for observation types AIRCRAFT, BRIGHT. TEMP. (satellite measurements) and GPSRO are suspiciously low and this is probably due to the fact that these observations indeed have correlated errors (thus, additional parameters describing these correlations should be included in the maximum-likelihood approach). In our study the corresponding coefficients will just be discarded for the tuning as in Desroziers et al. (2009).

4.2 Impact in terms of forecast performance

We have investigated the impact of tuning the background and observation variances in the global 4D-Var scheme. Figure 7 shows the differences between root-mean-square errors for the geopotential between the reference and the ex-

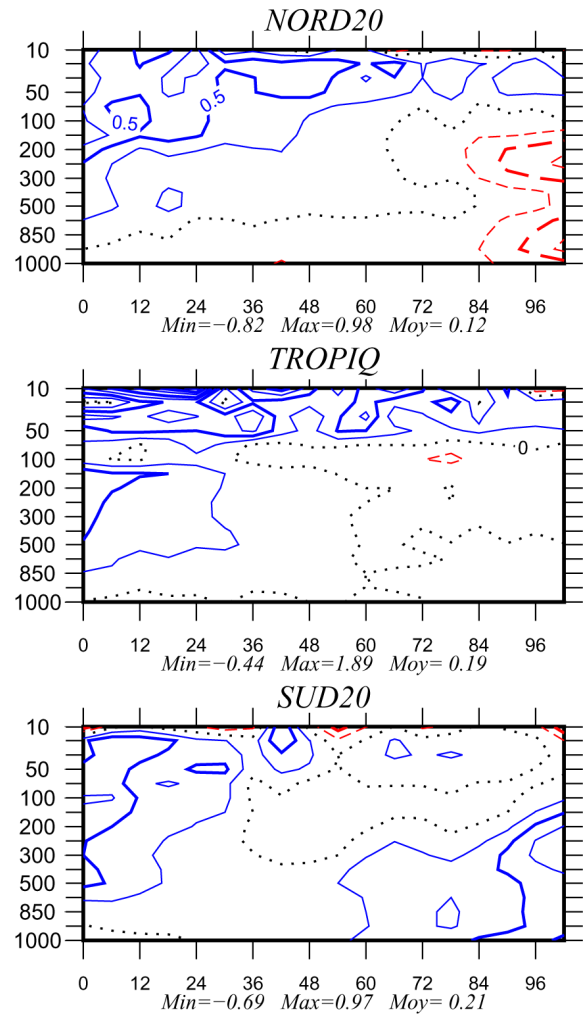


Fig. 7. Impact of tuning on forecast scores: differences between the root mean squared errors of the reference and of the experiment compared to independent ECMWF analyses (geopotential error in m), as a function of forecast range (in h). NORD20 is the domain on the globe that is north of 20° N, TROPIC is the domain between 20° N and 20° S, and SUD20 is south of 20° S. Blue (red) indicates better (worse) performance. The verification period is one month (September 2013).

periment (with tuned values of variances). The error is computed with respect to independent ECMWF analyses. The verification period is one month, which is generally sufficient to obtain reliable results. Bootstrap tests (not shown) indicate that the results are statistically significant during the first 36 h, with positive impact over the whole atmospheric column except for the surface. Other scores however point out to a degradation of the wind analyses in the Tropics with respect to the observations.

5 Conclusions

A posteriori diagnostics can help to detect misspecifications in the statistics of a data assimilation scheme. However, it is generally not possible to tell where the misspecification lies unless some further hypotheses are made (Talagrand, 2010).

Among those diagnostics, one is particularly simple. The value of the cost-function at the minimum should be close to its expected value $p/2$ where p is the number of observations (Bennett et al., 1993). It is possible to refine this diagnostic to sub-parts of the cost function (Talagrand, 1999). Despite care taken in the estimation of error covariance matrices, implementation of this diagnostic in early operational data assimilation schemes showed that the amplitude of the innovation vector was a priori overestimated by a factor 2–3 as reported by Talagrand (1999), possibly because of the misspecification of error correlations for satellite observations.

This paper introduces some additional results using the fact that the sub-parts of the cost function at the minimum can be expressed as quadratic forms of the innovation vector. For some specific distributions of the innovation vector, including the multivariate normal, expressions for the expectation, variances and cross-covariances can be given. In particular, the observation and background terms at the minimum are positively (generally weakly) correlated (when the innovation vector follows the expected statistics).

The expressions involve the trace of large matrices such as $\text{Tr}[(\mathbf{H}\mathbf{K})^2]$. It is shown that these matrices can be estimated from a randomized method (when \mathbf{K} is not explicitly available) by applying the assimilation scheme twice. The randomized method is shown to yield results that are consistent with the “simulated optimal innovation” method (Chapnik et al., 2006). Finally, it is advocated that the computation of $\text{Tr}[(\mathbf{H}\mathbf{K})^2]$ could be useful for variance tuning, in particular for the inflation factor of the background error covariance matrix. Further work will attempt to estimate additional covariance parameters following Purser and Parrish (2003) and also make use of recent advances in randomized linear algebra (Aune et al., 2012; Stein et al., 2013)

Acknowledgements. Part of this work was stimulated by a presentation given by M. Jarda at the International Conference on Ensemble Methods in Geophysical Sciences, Toulouse, November 2012. I am grateful to F. Rabier and to G. Desroziers for providing helpful comments on a previous version of this manuscript. This study benefited from the support of the RTRA STAE foundation within the framework of the FILAOS project.

Edited by: T. Gneiting

Reviewed by: R. Bannister and E. Aune



The publication of this article is financed by CNRS-INSU.

References

- Aune, E., Simpson, D., and Eidsvik, J.: Parameter estimation in high dimensional Gaussian distributions, *Stat. Comput.*, online first, 1–17, doi:10.1007/s11222-012-9368-y, 2012.
- Avron, H. and Toledo, S.: Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix, *Journal of the ACM*, 58, 8:1–8:34, 2011.
- Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. Roy. Meteor. Soc.*, 134, 1971–1996, 2008.
- Bekas, C., Kokiopoulou, E., and Saad, Y.: An estimator for the diagonal of a matrix, *Appl. Numer. Math.*, 57, 1214–1229, 2007.
- Belo-Pereira, M. and Berre, L.: The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model, *Mon. Weather Rev.*, 134, 2466–2489, 2006.
- Bennett, A., Leslie, L., Hagelberg, C., and Powers, P.: Tropical Cyclone Prediction Using a Barotropic Model Initialized by a Generalized Inverse Method, *Mon. Weather Rev.*, 121, 1714–1729, 1993.
- Bennett, A. F., Chua, B. S., Harrison, D. E., and McPhaden, M. J.: Generalized Inversion of Tropical Atmosphere–Ocean (TAO) Data and a Coupled Model of the Tropical Pacific. Part II: The 1995 La Nina and 1997 El Nino, *J. Climate*, 13, 2770–2785, 2000.
- Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Properties and first application of an error-statistics tuning method in variational assimilation, *Q. J. Roy. Meteor. Soc.*, 130, 2253–2275, 2004.
- Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, *Q. J. Roy. Meteor. Soc.*, 132, 543–565, 2006.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. Roy. Meteor. Soc.*, 120, 1367–1387, 1994.
- Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, *Q. J. Roy. Meteor. Soc.*, 124, 1783–1807, 1998.
- Daley, R.: Estimating observation error statistics for atmospheric data assimilation, *Annales Geophysicae*, 11, 634–647, 1993.
- Dee, D. and da Silva, A.: Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology, *Mon. Weather Rev.*, 127, 1822–1834, 1999.
- Dee, D. P.: On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation, *Mon. Weather Rev.*, 123, 1128–1145, 1995.
- Derber, J. and Bouttier, F.: A reformulation of the background error covariance in the ECMWF global data assimilation system, *Tellus A*, 51, 195–221, 1999.

- Desroziers, G. and Ivanov, S.: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation, *Q. J. Roy. Meteor. Soc.*, 127, 1433–1452, 2001.
- Desroziers, G., Berre, L., Chabot, V., and Chapnik, B.: A posteriori diagnostics in an ensemble of perturbed analyses, *Mon. Weather Rev.*, 137, 3420–3436, 2009.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Fisher, M.: Background error covariance modelling, in: ECMWF Seminar on recent developments in data assimilation for atmosphere and ocean, ECMWF, 45–63, 2003.
- Gauthier, P. and Thépaut, J.-N.: Impact of the digital filter as a weak constraint in the preoperational 4Dvar assimilation system of Météo-France, *Mon. Weather Rev.*, 129, 2089–2102, 2001.
- Gauthier, P., Charette, C., Fillion, L., Koclas, P., and Laroche, S.: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis, *Atmos. Ocean*, 37, 103–156, 1999.
- Genton, M. G., He, L., and Liu, X.: Moments of skew-normal random vectors and their quadratic forms, *Stat. Probabil. Lett.*, 51, 319–325, 2001.
- Girard, D.: Asymptotic optimality of the fast randomized versions of GCV and CL in ridge regression and regularization, *Ann. Stat.*, 19, 1950–1963, 1991.
- Guidard, V. and Fischer, C.: Introducing the coupling information in a limited-area variational assimilation, *Q. J. Roy. Meteor. Soc.*, 134, 723–735, 2008.
- Hollingsworth, A. and Lonnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus A*, 38, 111–136, 1986.
- Houtekamer, P. and Mitchell, H.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Hutchinson, M. F.: A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, *Commun. Stat.-Simul. C.*, 18, 1059–1076, 1989.
- Kucukkaraca, E. and Fisher, M.: Use of analysis ensembles in estimating flow-dependent background error variances, *Tech. Rep. 492*, ECMWF Technical Memorandum, 2006.
- Le Dimet, F. X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus A*, 38, 97–110, 1986.
- Lorenc, A. C.: Analysis methods for numerical weather prediction, *Q. J. Roy. Meteor. Soc.*, 112, 1177–1194, 1986.
- Lorenc, A. C.: Development of an operational variational assimilation scheme, *J. Meteorol. Soc. Jpn.*, 75, 339–346, 1997.
- Mathai, A. and Provost, S. B.: Quadratic forms in random variables: theory and applications, Marcel Dekker, Inc. New York, statistics, a series of textbooks and monographs, 1992.
- Mathai, A., Provost, S., and Hayakawa, T.: Bilinear Forms and Zonal Polynomials, Springer-Verlag, New York, Lect. Notes Stat., 102, doi:10.1007/978-1-4612-4242-0, 1995.
- Ménard, R. and Chang, L.-P.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: χ^2 -Validated Results and Analysis of Variance and Correlation Dynamics, *Mon. Weather Rev.*, 128, 2672–2686, 2000.
- Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation, *Mon. Weather Rev.*, 128, 2654–2671, 2000.
- Michel, Y.: Estimating deformations of random processes for correlation modelling: methodology and the one-dimensional case, *Q. J. Roy. Meteor. Soc.*, 139, 771–783, 2013.
- Muccino, J. C., Hubele, N. F., and Bennett, A. F.: Significance testing for variational assimilation, *Q. J. Roy. Meteor. Soc.*, 130, 1815–1838, 2004.
- Pailleux, J., Geleyn, J.-F., and Legrand, E.: La prévision numérique du temps avec les modèles Arpege et Aladin: Bilan et perspectives = Numerical weather prediction with the models Arpege and Aladin; assessment and prospects, *La Météorologie*, 32–60, 2000 in French.
- Parrish, D. and Derber, J.: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Purser, R. J. and Parrish, D. F.: A Bayesian technique for estimating continuously varying statistical parameters of a variational assimilation, *Meteorol. Atmos. Phys.*, 82, 209–226, 2003.
- Rabier, F.: Overview of global data assimilation developments in numerical weather-prediction centres, *Q. J. Roy. Meteor. Soc.*, 131, 3215–3233, 2005.
- Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Q. J. Roy. Meteor. Soc.*, 126, 1143–1170, 2000.
- Rasmussen, C. E. and Williams, C.: Gaussian Processes for Machine Learning, the MIT Press, 2006.
- Raynaud, L., Berre, L., and Desroziers, G.: Accounting for model error in the Météo-France ensemble data assimilation system, *Q. J. Roy. Meteor. Soc.*, 138, 249–262, 2012.
- Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd Edn., Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003.
- Sadiki, W. and Fischer, C.: A posteriori validation applied to the 3D-VAR Arpege and Aladin data assimilation systems, *Tellus A*, 57, 21–34, 2005.
- Simmons, A. and Hollingsworth, A.: Some aspects of the improvement in skill of Numerical Weather Prediction, *Q. J. Roy. Meteor. Soc.*, 128, 647–677, 2002.
- Stein, M. L., Chen, J., and Anitescu, M.: Stochastic Approximation of Score Functions for Gaussian Processes, *Ann. Appl. Stat.*, 7, 1162–1191, 2013.
- Talagrand, O.: A posteriori verification of analysis and assimilation algorithms, in: Workshop on diagnosis of data assimilation systems, 2–4 November 1998, ECMWF, Reading, UK, 1999.
- Talagrand, O.: Data Assimilation: Making Sense of Observations, in: Evaluation of Assimilation Algorithms, Springer-Verlag Berlin Heidelberg, 217–240, 2010.