Objective Validation and Evaluation of Data Assimilation

Olivier Talagrand

Laboratoire de Météorologie Dynamique, École Normale Supérieure
24, rue Lhomond, 75231 Paris cedex 05, France
talagrand@lmd.ens.fr

1. Introduction

These notes discuss some aspects of the objective validation and evaluation of algorithms for assimilation of meteorological observations. After a brief reminder of the basics of assimilation and of the theory of statistical linear estimation (Section 2), three classes of objective evaluation criteria are discussed and illustrated: assessment of assimilated fields against independent data and check of the consistency between *a priori* assumed and *a posteriori* observed probability distributions for the data errors (Section 3), objective check of optimality of a linear estimation scheme (Section 4). Parts of the contents of these notes have already been published in Talagrand (1999, 2003) and in Talagrand and Bouttier (2000). We also refer to Rodgers (2000) and Bennett (2002) for material of fundamental interest for questions discussed in these notes.

2. Basics on assimilation - the best linear unbiased estimator

The purpose of assimilation of meteorological observations is to estimate as accurately as possible the state of the atmospheric flow, using all available relevant information. The latter essentially consists of two parts

(a) The observations proper, which are distributed more or less regularly in both space and time, and vary in nature, accuracy, as well as in spatial and temporal resolution.

(b) The physical laws which govern the evolution of the flow, available in practice in the form of a discretised, and necessarily approximate, numerical model.

Either form of information is affected with some uncertainty, and that uncertainty must be taken into account in the assimilation process. That is neccessary for at least two reasons. First, a larger weight must be given, in one way or another, to more accurate data than to less accurate ones. And it is highly desirable to be able to quantify the uncertainty on the fields produced by the assimilation.

The most convenient way for describing uncertainty in a way that is both mathematically consistent and manageable is through probability distributions. This leads to describe the purpose of assimilation in terms of *probabilistic*, or *bayesian*, estimation. Stated in a few words, the ultimate purpose of assimilation is to determine the conditional probability distribution for the state of the atmospheric flow, given the data and the probability distribution describing the uncertainty affecting those data (see, *e.g.*, Lorenc 1986). However, if such a general formulation provides useful guidelines for the development of assimilation methods, it goes well beyond what can be achieved in practice. The probability distribution of the uncertainty affecting the various data is often very poorly known. And even if it was known, it would be totally impossible, with present means of computation, to practically describe a probability distribution in the state space of a Numerical Weather Prediction model, which can now have dimensions in the range 10^6-10^7.

One has to limit oneself to a much more modest goal. Most present assimilation algorithms are based, sometimes implicitly, on a linear approximation. A general description is as follows. Let us denote x the real, unknown, state of the atmospheric flow, expressed in a format appropriate for the purpose at hand (most often, in the format of the state vector of the numerical dynamical model used for the assimilation). The vector x belongs to *state space*, denoted $\mathcal{S}$, with dimension n. The available data are assumed to make up a vector z, belonging to *data space* $\mathcal{D}$, with dimension m. The link between the data and the unknown state vector is assumed to be of the form

$$z = \Gamma x + \zeta \tag{1}$$

where Γ is a known (mxn)-matrix, representing a linear operator from state space into data space, while ζ is a random vector in data space (the 'data error'), meant to represent the uncertainty on the data. We look for an estimate of x of the form

$$\boldsymbol{x}^a = \boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{z}$$

where the n-vector $\boldsymbol{\beta}$ and the (nxm)-matrix $\boldsymbol{A}$ are to be determined under the following two conditions

The estimate $\boldsymbol{x}^a$ is independent of the choice of the origin in state space (the result must be independent of whether temperatures are expressed in kelvins or celsius).

(a) The variance of the estimation error $\boldsymbol{x}^a - \boldsymbol{x}$ is minimum.

The solution to that problem is

$$\boldsymbol{x}^a = (\boldsymbol{\Gamma}^T \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^T \boldsymbol{S}^{-1} [\boldsymbol{z} - \boldsymbol{\mu}] \tag{2}$$

i.e.,

$$\boldsymbol{A} = (\boldsymbol{\Gamma}^T \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^T \boldsymbol{S}^{-1} \tag{3a}$$

.

$$\beta = -A\mu \tag{3b}$$

where $\boldsymbol{\mu} \equiv E(\boldsymbol{\zeta})$ and $\boldsymbol{S} \equiv E[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^T]$ are respectively the expectation and covariance matrix of the data error (here and in the following, E denotes mathematical expectation, and ^{T} vector or matrix transposition).

The corresponding estimation $\boldsymbol{x}^a - \boldsymbol{x}$ error is unbiased

$$E(\boldsymbol{x}^a - \boldsymbol{x}) = 0 \tag{4}$$

and has covariance matrix

$$\boldsymbol{P}^a \equiv E[(\boldsymbol{x}^a - \boldsymbol{x})(\boldsymbol{x}^a - \boldsymbol{x})^T] = (\boldsymbol{\Gamma}^T \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \tag{5}$$

The estimate $\boldsymbol{x}^a$ is the minimizer of the following scalar *objective function*, defined on state space

$$\mathcal{J}(\boldsymbol{\xi}) \equiv (1/2)\, [\boldsymbol{\Gamma}\boldsymbol{\xi} - (\boldsymbol{z}-\boldsymbol{\mu})]^T \boldsymbol{S}^{-1} [\boldsymbol{\Gamma}\boldsymbol{\xi} - (\boldsymbol{z}-\boldsymbol{\mu})] \tag{6}$$

The meaning of that expression is clear. For any vector $\boldsymbol{\xi}$ in state space, $\mathcal{J}(\boldsymbol{\xi})$ measures the misfit between the data and the analogue $\boldsymbol{\Gamma}\boldsymbol{\xi}$ of the data for state vector $\boldsymbol{\xi}$. The misfit is weighted by the inverse covariance $\boldsymbol{S}^{-1}$ of the data error, so as to give a large weight to accurate data, and a small weight to inacurrate data.

The estimate $\boldsymbol{x}^a$ is called the *Best Linear Unbiased Estimator* (*BLUE*) of $\boldsymbol{x}$ from $\boldsymbol{z}$. Its explicit determination requires, at least apparently, the *a priori* knowledge of the first- and second-order statistical moments (expectation and covariance matrix) of the data error $\boldsymbol{\zeta}$.

It is seen that the matrix $\boldsymbol{A}$ is a left-inverse of $\boldsymbol{\Gamma}$ ($\boldsymbol{A}\boldsymbol{\Gamma} = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the unit matrix of order n). This means that, when the data are exact ($\boldsymbol{\zeta}$=0), the estimated state $\boldsymbol{x}^a$ will be equal to the real state $\boldsymbol{x}$. That an estimation scheme at least does not degrade exact data is an obviously desirable quality. Conversely, any left-inverse of $\boldsymbol{\Gamma}$ is of form (3a), where $\boldsymbol{S}$ is a non-negative symmetric matrix.

Expressions (2) and (5) are easily verified to be invariant in any invertible linear change of coordinates, either in data or state space. This means for instance that the *BLUE* is independent of whether wind velocity is expressed in geometrical components, or in terms of vorticity and divergence, or of whether meteorological fields are defined by gridpoint values, or by spectral coefficients. The *BLUE* is also independent, for instance, of whether a vertical sounding is expressed as a temperature profile in function of pressure or, through integration of the hydrostatic equation, as a geopotential profile.

In particular, the objective function (6) is independent of the coordinates used in data space. The right-hand side of (6) defines a proper scalar product in data space, called the *Mahalanobis* scalar product associated with the covariance matrix $\boldsymbol{S}$.

The condition for the *BLUE* to be unambiguously defined is that the data matrix $\boldsymbol{\Gamma}$ be of rank n, rank$\boldsymbol{\Gamma} = n$ (the condition that $\boldsymbol{S}$ must be invertible, which seems to required by equation (2), is only apparent, and the *BLUE* remains unambiguoiusly defined when $\boldsymbol{S}$ is singular, and some of the data at least are assumed to be perfectly accurate). The meaning of the condition rank$\boldsymbol{\Gamma} = n$, which will be called the *determinacy condition*, is clear. It expresses that the data contain information, either directly or indirectly, on any component of the state vector $\boldsymbol{x}$. If the determinacy condition is not verified, some components of $\boldsymbol{x}$ remain undetermined. The determinacy condition implies that there must be at lesat as many individual scalar data as components of $\boldsymbol{x}$ to determined, $m \geq n$. We will set $m = n + p$.

In the case when the data error $\boldsymbol{\zeta}$ is gaussian, $\boldsymbol{\zeta} = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{S})$, the conditional probability distribution for $\boldsymbol{x}$, given the data, is the gaussian disribution $\mathrm{N}(\boldsymbol{x}^a, \boldsymbol{P}^a)$. Equations (2) and (5) therefore entirely solve the problem of bayesian estimation in the case of gaussian errors.

When the determinacy condition is verified (and independently of any gaussian hypothesis), it is always possible to transform the data, through an invertible change of origin and coordinates in data space, into the following form

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b \tag{7a}$$

$$\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon} \tag{7b}$$

The vector $\boldsymbol{x}^b$, which has dimension n, is an explicit estimate of the unknown state vector $\boldsymbol{x}$. It will be called the *background estimate* of $\boldsymbol{x}$ (although that denomination must not be understood as implying that the knowledge of $\boldsymbol{x}^b$ is necessarily anterior in any way to the knowledge of the other data). The vector $\boldsymbol{y}$, which has dimension p, is an additional set of data, linked to the real state vector $\boldsymbol{x}$ through the (pxn)-matrix $\boldsymbol{H}$. The errors $\boldsymbol{\zeta}^b$ and $\boldsymbol{\varepsilon}$ can be assumed, without loss of generality, to have zero expectation

$$E(\boldsymbol{\zeta}^b) = 0 \quad ; \quad E(\boldsymbol{\varepsilon}) = 0 \tag{8}$$

As for the corresponding covariance matrices, they will be denoted:

$$E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{b\mathrm{T}}) = \boldsymbol{P}^b \quad ; \quad E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\mathrm{T}) = \boldsymbol{R} \tag{9}$$

(note that the matrix $E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{b\mathrm{T}})$ is often denoted $\boldsymbol{B}$). It can be further assumed, again without loss of generality, that the errors $\boldsymbol{\zeta}^b$ and $\boldsymbol{\varepsilon}$ are mutually uncorrelated

$$E(\boldsymbol{\zeta}^b \boldsymbol{\varepsilon}^\mathrm{T}) = 0 \tag{10}$$

Equations (7-8-10) are a fairly good description of the actual conditions of most meteorological and oceanographical applications. A background $\boldsymbol{x}^b$ is usually available, in the form of a climatological estimate, or of a forecast coming from the past. As for the additional vector of data $\boldsymbol{y}$, it usually consists, for the most part at least, of observations, which may be either synchronous or distributed over a period of time. For that reason, and for lack of a better expression, the vector $\boldsymbol{y}$ will be called in the sequel the *observation vector*, belonging to *observation space* O, with dimension p.

Under conditions (8-9-10), the expressions (2) and (5) for the *BLUE* and the associated error covariance matrix assume the familiar forms

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b) \tag{11a}$$

$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1} \boldsymbol{H}\boldsymbol{P}^b \tag{11b}$$

Both the estimate $\boldsymbol{x}^a$ and the associated error covariance matrix $\boldsymbol{P}^a$ are expressed in those equation as the sum of the background quantities $\boldsymbol{x}^b$ and $\boldsymbol{P}^b$, and of an appropriate correction. The correction on $\boldsymbol{x}^b$ is proportional, through the *gain matrix* $\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1}$, to the so-called *innovation vector*

$$\mathbf{d} \equiv \mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{b}} = \boldsymbol{\varepsilon} - \mathbf{H}\boldsymbol{\zeta}^{\mathrm{b}} \tag{12}$$

The innovation vector is the discrepancy between the observation vector $\boldsymbol{y}$ and the analogue of the observated quantities in the background. It is of course only to the extent that the discrepancy is different from zero that the background has to be corrected. The innovation is a function of the background and observation errors only, and is independent of the state vector $\boldsymbol{x}$. We can note that the matrix $\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}$, whose inverse appears in the gain matrix, is the covariance matrix of the innovation

$$\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R} = E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}) \tag{13}$$

Equations (11) are exactly equivalent to equations (2) and (5). Equations (2) and (5) are often (but not always) more convenient for theoretical developments, while equations (11) are usually more useful for practical numerical implementation.

Most analysis and assimilation algorithms that have been used so far, or are used now, in meteorological or oceanographical applications are particular examples of the *BLUE*. Optimal Interpolation, Kalman filtering in its various simplified and/or extended forms, Kalman smoothing, three-dimensioanl variational analysis, in both its primal and dual formulations, four-dimensional variational assimilation, either of the strong- or weak-constraint type, and in both its primal and dual formulations, all those algorithms can be described as particular applications of the general approach that has just been described. Innumerable variants exist as to the choice of the data, the definition of the data operator Γ, the *a priori* specification of the first- and second-order statistical moments of the errors affecting the data, and the numerical algorithms that are used for actually carrying out the necessary computations. But the basic purpose is always to obtain a variance-minimizing linear (or quasi-linear) combination of the data. The only real exception so far is *ensemble assimilation*, which does not produce one estimate of the state of the flow plus a measure of the associated uncertainty, but produces rather an ensemble of estimates which are supposed to sample the conditional probability distribution of the state of the flow (see, *e. g.,* Evensen and van Leeuwen 2000, or Houtekamer and Mitchell 2001). And still, most algorithms for ensemble assimilation are partially linear in that the formula (11a) is used for updating a predicted ensemble with new observations. Distinctly non-linear methods are also used for *quality control* of observations, *i.e.* for detection of erroneous observations (see, *e.g*., Lorenc, 1997). But, as important as quality control is for NWP, it is a very limited and specific aspect of the whole process of assimilation.

3. Objective Validation and Evaluation

If the theory of assimilation algorithms is now well understood, the same cannot be said of *a posteriori* validation and evaluation of existing algorithms. The only way to objectively assess the quality of the fields produced by an assimilation algorithm is by comparison with independent unbiased data, *i. e.*, with data that are affected by errors which, in addition to being zero on statistical average, are statistically independent of the errors affecting the data that have been used in the assimilation. Those conditions can be difficult to achieve, and are in any case impossible to objectively assess on the basis of the data themselves. The problem is complicated by the presence in the data of representativeness errors, due to the different spatial and temporal resolutions of the data and the assimilating model. Representativeness errors can be correlated, even if the instrumental errors of the data are not correlated. And of course, comparison with independent data can measure the quality of the assimilation only for those parameters for which independent data are available. In addition, if it allows to compare for instance the relative quality of two different assimilation procedures, it says nothing as to the possibly optimal character of those procedures.

The determination of the *BLUE* requires the *a priori* specification of the first- and second-order statistical moments of the data errors. How is is possible to objectively determine those required quantities ? The only objective source of information on the data errors consists of the combinations of the data that are algebraically independent of the state vector ***x***. When the data are written in format (7), those independent combinations make up the innovation vector (12), which thus contains all the objective information on the data errors. How can statistics of the innovation vector be used to determine the statistical moments whose specification is required for determining the *BLUE* ? To answer that question, we consider an estimation scheme of the form

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{K}\ (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b) \tag{14}$$

where the gain matrix ***K*** can now be any (pxn)-matrix, not necessarily of the optimal form (11a) (one question of interest is precisely whether a known gain matrix is optimal or not). This is equivalent to saying that the estimation scheme produces the exact state vector ***x*** when implemented on exact data. Or that the scheme is of form (2), where the matrix ***S*** is non-negative (but not neceessarily optimal).

Consider the difference

$$\delta \equiv \boldsymbol{z} - \boldsymbol{\Gamma}\boldsymbol{x}^{\mathrm{a}} \tag{15}$$

i.e. the *a posteriori* difference between the data vector and the analogue of the data in the analysed fields. That vector will be called the Data-minus-Analysis (briefly, DmA) difference. In formulation (7), it reads

$$\delta = \begin{pmatrix} \boldsymbol{x}^b - \boldsymbol{x}^a \\ \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^a \end{pmatrix} = \begin{pmatrix} -\boldsymbol{K}\boldsymbol{d} \\ (\boldsymbol{I}_p - \boldsymbol{H}\boldsymbol{K})\boldsymbol{d} \end{pmatrix}$$

where $\boldsymbol{I}_p$ is the unit matrix of order p. For given gain matrix ***K***, the DmA difference is a linear invertible function of the innovation vector ***d***. It is therefore exactly equivalent to perform statistics on either one of those two vectors.

The estimated state $\boldsymbol{x}^{\mathrm{a}}$ minimizes the objective function (6). The meaning of the minimization is clear. $\boldsymbol{\Gamma}\boldsymbol{x}^{\mathrm{a}}$ is the point in the image space $\Gamma(\mathrm{S})$ which lies closest, in the sense of the ***S***-Mahalanobis scalar product, to the data vector ***z***. $\boldsymbol{\Gamma}\boldsymbol{x}^{\mathrm{a}}$ is therefore the orthogonal projection, in the sense of that scalar product, of ***z*** onto $\Gamma(\mathrm{S})$. This suggests to decompose the data space D into the image space $\Gamma(\mathrm{S})$ (which, because of the determinacy condition, has dimension n) and the space $\perp\Gamma(\mathrm{S})$ orthogonal to $\Gamma(\mathrm{S})$ according to the ***S***-Mahalanobis scalar product (which has dimension p). In that decomposition, the covariance matrix ***S*** reads

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_1 & 0 \\ 0 & \boldsymbol{S}_2 \end{pmatrix} \tag{16a}$$

where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are positive definite matrices with respective dimensions nxn and pxp. The data matrix Γ reads

$$\Gamma = \begin{pmatrix} \Gamma_1 \\ 0 \end{pmatrix} \tag{16b}$$

where Γ_1 is an (nxn)-matrix which, because of the determinacy condition, is invertible. Denoting by ζ_1 and ζ_2 the projections of the error vector ζ onto $\Gamma(S)$ and $\perp\Gamma(S)$ respectively, the components of the data vector $\boldsymbol{z}$ read

$$z_1 = \Gamma_1 \boldsymbol{x} + \zeta_1$$

$$z_2 = \zeta_2$$

For the sake of generality, we now assume that the error vector may have non zero expectation. The projection onto $\Gamma(S)$ of the unbiased data vector is equal to $z_1 - E(\zeta_1)$. Equations (2), (5) and (16) then show that the *BLUE* $\boldsymbol{x}^a$ and the associated estmation error covariance matrix are respectively equal to

$$\boldsymbol{x}^a = \Gamma_1^{-1} [z_1 - E(\zeta_1)] \tag{17a}$$

and

$$\boldsymbol{P}^a = \Gamma_1^{-1} \boldsymbol{S}_1 \Gamma_1^{-T} \tag{17b}$$

The DmA difference (15) is orthogonal to $\Gamma(S)$ and, considered as a vector of $\perp\Gamma(S)$, equal to ζ_2. It therefore has expectation $E(\zeta_2)$ and covariance matrix $\boldsymbol{S}_2$. These quantities are independent of the quantities $E(\zeta_1)$ and $\boldsymbol{S}_1$ which determine the *BLUE* and the associated error (equations 17). This shows that any expectation and covariance matrix of the DmA difference are compatible with any estimation scheme of form (14). The knowledge of the statistics of the DmA difference, or equivalently of the innovation vector, is totally useless for determining the *BLUE* or the associated error covariance matrix. Consistency between the observed and specified statistics of the innovation vector is neither a sufficient, nor even a necessary, condition for optimality of the estimate.

This is true of course in the absence of any knowledge on the data error other than the innovation vector. As a simple example, let us consider the case when two observations of a scalar quantity x are available, of the form

$$z_1 = x + \zeta_1$$
$$z_2 = x + \zeta_2$$

The only combination of those observations that is independent of x is the difference $z_1 - z_2 = \zeta_1 - \zeta_2$ (*i. e.*, the innovation vector if z_2 is arbitrarily considered as being the 'background'). The general result that has just been stated is that statistics of $z_1 - z_2$, obtained for instance from a time series of independent couples of simultaneous observations, are totally useless for estimating x. But if it is independently known (or if it be can reasonably assumed) that the observations z_1 and z_2 are unbiased $[E(\zeta_1) = E(\zeta_2) = 0]$ and of same statistical quality $[E(\zeta_1^2) = E(\zeta_2^2)]$, then the optimal estimate is necessarily $x^a = (1/2)\ (z_1 + z_2)$. However those properties cannot be inferred from statistics of $z_1 - z_2$ [strictly speaking, the hypothesis that both observations are unbiased would be unvalidated if it turned out that $E(z_1 - z_2) \neq 0$; but the more general hypothesis that $E(\zeta_1 + \zeta_2) = 0$, which also leads to $x^a = (1/2)\ (z_1 + z_2)$, cannot be checked against the statistics of $z_1 - z_2$].

One can wonder why statistical consistency between the *a priori* assumed and *a posteriori* observed statistics of the innovation vector is not of critical importance. The fundamental reason is that, contrary to what a cursory look at, say, equation (6) might lead to think, it is not necessary, in order to determine the

BLUE $\boldsymbol{x}^a$ and the associated estimation error covariance matrix $\boldsymbol{P}^a$, to know entirely the expectation μ and the covariance matrix $\boldsymbol{S}$ of the data error. Equations (2) and (5) show that it is sufficient to know respectively $\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1}\mu$ (which has dimension *n*, instead of *m* for μ) and $\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1}$ (which has dimension *n*x*m*, instead of *m*x*m* for $\boldsymbol{S}^{-1}$). The significance of that remark becomes clear in the decomposition (16). In order to determine the *BLUE* and the associated error, it is not necessary to know entirely the $\boldsymbol{S}$-Mahalanobis scalar product. It is sufficient to be able to identify the space $\perp\Gamma(\mathsf{S})$, and to know the $\boldsymbol{S}$-metric in the space $\Gamma(\mathsf{S})$. The metric in $\perp\Gamma(\mathsf{S})$ is useless. Similarly, concerning the expectation μ, only its $\boldsymbol{S}$-projection onto the space $\Gamma(\mathsf{S})$ is required. Any inconsistency between the *a priori* assumed and *a posteriori* observed statistics of the innovation vector can always be mathematically explained out by a misspecification of the additional, useless, degrees of freedom.

The fact that it is not necessary to know entirely μ and $\boldsymbol{S}$ in order to determine the *BLUE* and the associated uncertainty, if it is important from a theoretical point of view, is however of little practical importance, and cannot be taken davantage of for reducing the number of quantities which must be *a priori* specified for performing the assimilation. The $\boldsymbol{S}$-metric depend on the statistical properties of the errors affecting the various data, and it not possible to identify *a priori*, without knowing those errors, which parameters will be useful, and which will not. In addition, the data constantly vary over time in type, number, and spatio-temporal distribution, and one particular parameter which may be useless one day may be required the following day.

The general conclusion is that, if one wants to use the statistics of the innovation vector, or of any quantity derived from the innovation vector, for drawing inferences that can be useful for the assimilation, one must necessarily make independent hypotheses. Those independent hypotheses cannot be objectively validated, at least not on the basis of the innovation vector itself. Such hypotheses are actually very commonly done (and very often implicitly). For instance, a systematic bias in the innovation vector, at least in the components of the innovation vector obtained by comparison with well-calibrated observations, is usually interpreted as resulting from a bias in the background. For another example, several authors (Hollingsworth and Lönnberg, 1986, Daley, 1993) have studied the horizontal correlation of the innovation vector obtained from radiosonde observations. It is reasonable to assume that radiosonde observation errors are horizontally uncorrelated. In these conditions, a direct estimate of the horizontal correlation of the forecast error can be obtained. In addition, if the observation and forecast errors are supposed to be uncorrelated, the residual obtained by extrapolating the covariance to zero horizontal distance provides an estimate of observation error variance. One conclusion from these studies is that the 6-hour forecast error is typically of the same magnitude as the observational error.

The number of statistical diagnostics that can in principle be implemented on either the innovation or the DmA vector is in practice unlimited. Critical aspects are the statistical significance of the diagnostics on the one hand, and whether or not independent appropriate information is available for usefully exploiting any observed inconsistency between the *a priori* assumed and *a posteriori* observed statistics. Many *adaptive schemes* have been defined for progressively adjusting the expectation or covariance parameters on the basis of observed statistical inconsistencies. In the context of Kalman filtering, and independently of meteorological or oceanographical applications, one can mention the early works of Mehra (1972) and Godbole (1974). For more recent meteorological and oceanographical applications, see, *e. g*., Blanchet *et al*. (1997) or Dee *et al*. (1999).

Keeping in mind that a lack of consistency between *a priori* assumed and *a posteriori* observed statistics does not constitute in itself a proof of non-optimality, we now briefly describe a number of basic diagnostics. In a consistent system, the innovation and the DmA vectors have zero expectation. If they are observed to have non-zero expectation, it necessarily means that a systematic bias in the data has not been properly taken into account. This fact has been systematically exploited by Dee and Da Silva (1998) who, assuming any bias to come from the background, have developed algorithms for constantly correcting the latter.

Using equation (2), the covariance matrix of the DmA difference is shown to be equal to

$$\mathrm{E}(\delta\delta^{\mathrm{T}}) \ = \ \mathbf{S} - \mathbf{\Gamma P}^{\mathrm{a}}\mathbf{\Gamma}^{\mathrm{T}}$$

The term subtracted on the right-hand side is positive definite [it is actually the covariance matrix of the vector $\boldsymbol{\Gamma}(\boldsymbol{x}^a\text{-}\boldsymbol{x})$]. This means that the variance of any component of the DmA difference is less than the variance of the corresponding component of the error. Optimally assimilated fields statistically fit the data to within the accuracy of the latter (Hollingsworth and Lönnberg 1989, have called *efficient* an assimilation system that possesses that particular property). If an unbiased assimilation system does not fit the data to within their assumed accuracy, that means that the error covariance matrix $\boldsymbol{S}$ has been misspecified (the misspecification may of course be in the variance of the error affecting the ill-fitted data).

The objective function (6) takes at its minimum $\boldsymbol{x}^a$ the value (assuming observations have been unbiased)

$$\mathcal{J}(\boldsymbol{x}^a) \ = \ (1/2)\ [\boldsymbol{\Gamma x}^a - \boldsymbol{z}]^{\mathrm{T}}\ \boldsymbol{S}^{-1}\ [\boldsymbol{\Gamma x}^a - \boldsymbol{z}]$$

It is the squared $\boldsymbol{S}$-Mahalanobis norm of the DmA difference. Using the background-observation decomposition (7), standard matrix manipulations lead to

$$\mathcal{J}(\boldsymbol{x}^a) \ = \ (1/2)\ \boldsymbol{d}^{\mathrm{T}}\ [\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1}\ \boldsymbol{d} \qquad (18)$$

As already mentioned, the matrix $\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}$ is the covariance matrix $E(\boldsymbol{dd}^{\mathrm{T}})$ of the innovation $\boldsymbol{d}$, and $\mathcal{J}(\boldsymbol{x}^a)$ is therefore the squared Mahalanobis norm of $\boldsymbol{d}$ with respect to its own covariance matrix. Writing equation (18) in the basis of the principal components of $\boldsymbol{d}$, in which $E(\boldsymbol{dd}^{\mathrm{T}})$ is the unit matrix of order p, it is easily seen that, on statistical expectation

$$E[\mathcal{J}(\boldsymbol{x}^a)] \ = \ p/2 \qquad (19)$$

If in addition the data errors are gaussian, $\mathcal{J}(\boldsymbol{x}^a)$ can be shown to follow a χ^2 probability distribution of order p, which means that its standard deviation is equal to $\sqrt{(p/2)}$. For large values of p, the distribution of $\mathcal{J}(\boldsymbol{x}^a)$ must therefore be strongly concentrated about its expectation. This is likely to remain true, even if the data errors are not strictly gaussian, as long as they have a ‘reasonably’ symmetric unimodal distribution.

Equation (19) provides a very simple diagnostic of the overall global consistency of an assimilation algorithm. A number of authors (Ménard and Chang, 2000, Talagrand and Bouttier, 2000, Cañizares *et al.*, 2001, see also Bennett, 2002) have implemented this diagnostic on various systems. Tests have been performed on the operational 4D-variational assimilation system of the European Centre for Medium-range Weather Forecasts for January 2003. The dimension of the state vector of the minimization is about $n \approx 8\text{x}10^6$, while the number p of observations fluctuates about $1.4\text{x}10^6$ (the assimilation window is 12 hours). The ratio $2\mathcal{J}(\boldsymbol{x}^a)/p$, which should statistically be equal to 1, turns out to fluctuate in the range 0.40-0.45. This means that the covariance matrix $E(\boldsymbol{dd}^{\mathrm{T}})$ is largely overestimated by the assimilation system. When excluding satellite observations from the assimilation (which reduces the number of observations to about 2-$3\text{x}10^5$), the ratio $2\mathcal{J}(\boldsymbol{x}^a)/p$ takes values in the range 1.-1.05, much closer to overall consistency. It is therefore reasonable to assume (although, according to the remark made above, it cannot be rigourously proved) that the overestimation of $E(\boldsymbol{dd}^{\mathrm{T}})$ is due to misspecification of the covariance matrix of the errors in satellite observations. In the ECMWF system (as in most other meteorological assimilation systems), errors in satellite observations are assumed to be spatially uncorrelated. This is unlikely to be true, at least for observations performed by a same instrument, and other tests (Thépaut *pers. com.*) suggest the errors are correlated over distances of a few hundred kilometres. Neglecting spatial correlation would lead to giving too large a weight to satellite observations in the analysis. In order to compensate for that effect, an expedient solution has been to artificially increase the assumed variance of the errors. This probably explains the significant overestimation of $E(\boldsymbol{dd}^{\mathrm{T}})$ in the ECMWF system.

Similar results, with similar conclusions, have been obtained by other authors for other meteorological assimilation systems (Sadiki and Fischer 2000, Payne *pers. com.*). A different series of diagnotics has recently been performed by Weaver *et al.* (2003), who have applied criterion (19) to a system of variational assimilation of oceanographical observations. They performed both three- and four-dimensional assimilation. In the three-dimensional system, the ratio $2\mathcal{J}(\boldsymbol{x}^a)/p$ is on average equal to 0.9, but varies in the range 0.7-1.3. In the four-dimensional system, $2\mathcal{J}(\boldsymbol{x}^a)/p$ fluctuates between 0.6 and 0.9. The authors attribute the larger inconsistency of the four-dimensional system to the fact that the background error (at the beginning of each successive assimilation cycle) is kept at the same value from one assimilation cycle to the next, and does not take into account the possibility that the accuracy of the backgrounds is likely to be improved by the successive assimilations. As for the large fluctuations of the ratio $2\mathcal{J}(\boldsymbol{x}^a)/p$, the authors explain it by significant variations of the actual error in the background, with the result that the pre-specified background error covariance matrix may be correct on average (at least in the three-dimensional system), but not in individual situations.

Diagnostics that are similar to the global diagnostic (19), but also more refined, can be defined. The objective function (6) will most often be the sum of a number of independent terms, *viz.*,

$$\mathcal{J}(\xi) \;=\; \Sigma_{k=1,\ldots,K}\, \mathcal{J}_K(\xi)$$

where

$$\mathcal{J}_k(\xi) \;\equiv\; (1/2)\,(\Gamma_k\xi - z_k)^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}(\Gamma_k\xi - z_k)$$

In this equation (where again data are assumed to be unbiased), z_k is an m_k–dimensional component of the data vector z ($\Sigma_k m_k = m$), and the rest of the notation is obvious. The inverse estimation error covariance matrix is easily obtained from equation (5) as

$$[\boldsymbol{P}^a]^{-1} \;=\; \Sigma_k\, \boldsymbol{\Gamma}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{\Gamma}_k$$

Left-multiplying by $\boldsymbol{P}^a$, and then taking the trace of the result, yield

$$\begin{aligned} 1 &= (1/n)\, \Sigma_k\, \mathrm{tr}(\boldsymbol{P}^a \boldsymbol{\Gamma}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{\Gamma}_k) \\ &= (1/n)\, \Sigma_k\, \mathrm{tr}(\boldsymbol{S}_k^{-1/2}\, \boldsymbol{\Gamma}_k\, \boldsymbol{P}^a\, \boldsymbol{\Gamma}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1/2}) \end{aligned} \qquad (20)$$

where use has been made, for obtaining the last equality, of the fact that the trace of the product of two matrices is not modified when the order of the factors is reversed. This expression shows that the quantity $(1/n)\ \mathrm{tr}(\boldsymbol{S}_k^{-1/2}\, \boldsymbol{\Gamma}_k\, \boldsymbol{P}^a\, \boldsymbol{\Gamma}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1/2})$ is a measure of the relative contribution of the subset of data z_k to the overall accuracy of the analysis. Equation (20) (like actually all equations in these notes) is valid in any system of coordinates in data and state spaces, and the measure $(1/n)\, \Sigma_k\ \mathrm{tr}(\boldsymbol{S}_k^{-1/2}\, \boldsymbol{\Gamma}_k\, \boldsymbol{P}^a\, \boldsymbol{\Gamma}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1/2})$ is absolutely intrinsic. In particular, given any subset z' of the data vector, its contribution to the objective function is independent of whether, in the coordinates used in data space, the errors affecting z' are correlated or not with the errors affecting the other components of z. Equation (20) is therefore valid, and can be used as a diagnostic tool, for any subset of data in any basis in data space. This defines a powerful and consistent tool, for instance in so-called Observing System Simulation Experiments, intended at estimating the usefulness of hypothetical systems of observations (for a similar, but somewhat different diagnostic, see Fisher 2003).

It can be further shown (Talagrand, 1999, Desroziers and Ivanov, 2001) that the expectation of the term $\mathcal{J}_k(\xi)$ at the minimum of the objective function is equal to

$$E[J_k(\boldsymbol{x}^a)] \;=\; (1/2)\,[m_k \;-\; \mathrm{tr}(\boldsymbol{S}_k^{-1/2}\,\boldsymbol{H}_k\,\boldsymbol{P}^a\boldsymbol{H}_k^{\mathrm{T}}\,\boldsymbol{S}_k^{-1/2})] \tag{21}$$

where the same trace appears on the right-hand-side as in equation (20). Equation (21) includes equation (19) as a particular case. And it provides the basis for further evaluation of the consistency of an assimilation scheme. It suffices to compare the trace of $\boldsymbol{S}_k^{-1/2}\,\boldsymbol{H}_k\,\boldsymbol{P}^a\,\boldsymbol{H}_k^{\mathrm{T}}\,\boldsymbol{S}_k^{-1/2}$, as computed directly ans as determined statistically, through equation (21), from results of assimilations. The matrix $\boldsymbol{S}_k^{-1/2}\,\boldsymbol{H}_k\,\boldsymbol{P}^a\,\boldsymbol{H}_k^{\mathrm{T}}\,\boldsymbol{S}_k^{-1/2}$ is symmetric non-negative, of order m_k. Even for large values of m_k, it can be numerically computed, at an acceptable cost, through techniques originally developed for generalized cross-validation (Wahba *et al.* 1995, Desroziers and Ivanov 2001, Fisher 2003). Desroziers and Ivanov (2001) have shown that, if the observation error is supposed to be uncorrelated in space and uncorrelated with the background error, equation (21) can be used for estimating the observation and background error variances. This is in essence a systematic extension of the already mentioned works by Hollingsworth and Lönnberg (1986), and Daley (1993). Along the same lines, Chapnik *et al.* (2003) have used equation (21) to tune the variances of the observational errors in the various channels of the TOVS instrument, carried by the satellites of the NOAA series. This has led to a significant change (by a factor 8) of one variance.

4. Checking Optimality

The various diagnostics that have been presented in the previous sections allow objective comparison of the quality of different assimilation schemes, or evaluation of the internal consistency of a given scheme. They say nothing as to the optimality, or otherwise, of a given scheme. The *BLUE* is defined on conditions of statistical unbiasedness and minimum estimation error variance. As a consequence, the estimation error $\boldsymbol{x}^a - \boldsymbol{x}$, in addition to being unbiased, must be statistically uncorrelated with the DmA difference or, equivalently, with the innovation vector. This is expressed by equation (11a), where the second term on the right-hand side is the orthogonal projection, in the sense of covariance, of (minus) the background error $\boldsymbol{x} - \boldsymbol{x}^b$ onto the space spanned by the innovation $\boldsymbol{y} - \boldsymbol{Hx}^b$. The optimality condition is often expressed, in an exactly equivalent way, by saying that a sequential algorithm for assimilation is optimal if, and only if, the temporal sequence of innovation vectors is unbiased and uncorrelated (Kailath, 1968).

This optimality condition can be objectively checked against independent observations. Let us consider an observation of the form

$$\mathbf{w} = \mathbf{Dx} + \gamma$$

where $\boldsymbol{D}$ is a known linear operator, and the error γ is assumed to be unbiased and uncorrelated with the data error ζ, and therefore with the innovation $\boldsymbol{d}$. Optimality of the estimate $\boldsymbol{w}^a = \boldsymbol{Dx}^a$ of $\boldsymbol{w}$ is equivalent to the conditions that it be statistically unbiased

$$E(\boldsymbol{w} - \boldsymbol{Dx}^a) = 0 \tag{22}$$

and uncorrelated with the innovation

$$E[(\boldsymbol{w} - \boldsymbol{Dx}^a)\boldsymbol{d}^{\mathrm{T}}] = 0 \tag{23}$$

If the unbiasedness condition (S) is usually checked in assimilation systems, the uncorrelatedness condition (S), in spite of its simplicity, has so far been rarely used. One of the few examples is a work by Daley (1992), who computed the correlation of the innovation sequence for the sequential assimilation system that was then in use at the Canadian Meteorological Centre (that system is described by Mitchell *et al.* 1990). Daley found significantly non-zero correlations, reaching values of more than 0.4 for the 500-hPa geopotential innovation, at time-lag 12 hours. Similar tests, performed more recently on a system for assimilation of oceanographical observations, led to correlation values around 0.3 (Miller *pers. com.*). It would certainly be very instructive to systematically implement diagnostic (23) on assimilation systems, and especially of course on operational systems.

5. Conclusions

We have presented, and briefly discussed and illustrated on a few examples, three classes of diagnostics for validation and evaluation of assimilation algorithms.

The first class consists simply in comparison of the fields produced by the assimilation with independent data. It provides of course the only objective measure of the quality of an assimilation algorithm, and the only way to compare the quality of different algorithms. On the other hand, it says nothing as to the possibly optimal character, in any sense, of an algorithm.

The second class of diagnostics evaluates the internal consistency of an assimilation algorithm, *i. e.* the consistency between the *a priori* assumed probability distribution (strictly speaking, expectation and covariance matrix) of the data error, and the real distribution. This comparison can be done only to the extent the real distribution of the error can be known. All the objective information about the data error contained in the data themselves lies in the overdeterminacy of those data. This information is entirely described by the observed probability distribution of the innovation vector (or of any quantity, such as the data-minus-analysis difference, that is a function of the innovation). This second class of diagnostics therefore always reduces to a comparison, either direct or indirect, between the *a priori* assumed and the *a posteriori* observed probability distributions of the innovation. However, redundant parameters are present in the *a priori* specification of the probability distribution, and in any observed inconsistency can always be mathematically attributed to a misspecification of those redundant parameters, without consequences for the result of the assimilation, nor even for the corresponding estimated estimation error. This second class of diagnostics can therefore be useful only if independent hypotheses can be made about the probability distribution of the data error. Such independent hypotheses have of course always been made (often implicitly) since the very beginnings of the development of assimilation techniques. Systematic use of statistically reliable diagnostics of the innovation vector, which must go together with a critical assessment of independent hypotheses, will continue to be extremely useful. Subjective judgment, aided by experience, will always be fundamental here.

The third class of diagnostics is intended at checking the optimality of a linear least-error-variance assimilation scheme. For such a scheme, optimality is equivalent to the condition that the estimation error must be statistically unbiased, and uncorrelated with the innovation. It can be objectively checked provided independent data are available. For some resaon, this type of diagnostic has been rarely used so far, and it would certainly be very useful to systematically implement it on linear assimilation schemes.

All the developments in these notes have been restricted to linear statistical estimation, which leads to the *BLUE* defined by equations (2) and (5). It may be useful to briefly discuss whether, and how, the above diagnostics can be used in the more general context of non-linear bayesian estimation. The first class of diagnostics, which only compares estimated fields with independent data, has of course nothing to do with the estimation process in itself, and can be used as such for evaluating and comparing estimation methods of any kind.

Concerning the second class, the fact that objective information about the data errors lies in the overdeterminacy of the data is always true. To the extent the data overdetermine the unknown state of the system, it will always be possible to extract from those data, by appropriate algebraic elimination, the analogue of the linear innovation, *i. e.* quantities that depend only on the data error, and not on the unknown state. And it will always be possible to compare the *a priori* assumed and the *a posteriori* observed probability distributions of that generalized innovation. It is not absolutely clear on the other hand whether it will always be possible to 'explain out' an observed inconsistency by attributing it to misspecification of redundant parameters in the data error probability distribution. The author does not know of any rigourous proof of that fact, but it seems to entirely result from the numerical dimensions involved (m and n), and not from the linearity of the estimation scheme (even though the proof given above is linear).

The third class of diagnostics, on the other hand, which is based on minimization of the variance of a linear combination of the data, *i. e.* on an orthogonal projection, is linear by essence. It is of course a check of optimality of the estimation in the case the data error is known *a priori* to be gaussian, but this is so because gaussianity as such implies linearity. And even in the gaussian case, the orthogonality criterion provides a check of the optimality of the estimate $\boldsymbol{x}^a$, not of the correctness of the corresponding estimated estimation error covariance matrix $\boldsymbol{P}^a$. More generally, the conditional probability distribution of the state vector, given the data, depends on the probability distribution of the data error. Unless it is possible to objectively check that the latter has been correctly specified (and it is difficult to imagine how that could be done), it will never be possible to check the correctness of the probability distribution produced by the estimation process.

References

Bennett, A. F. (2002) *Inverse Modeling of the Ocean and Atmosphere*. Cambridge, United Kingdom: Cambridge University Press.

Blanchet, I., C. Frankignoul and M. A. Cane (1997) A Comparison of Adaptive Kalman Filters for a Tropical Pacific Ocean Model. *Mon. Wea. Rev., 125*, 40-58.

Cañizares, R., A. Kaplan, M. A. Cane and D. Chen, S. E. Zebiak (2001) Use of data assimilation via linear low-order models for the initialization of El Niño - Southern Oscillation predictions. *J. Geophys. Res. 106 (C12)*, 30,947-30,959.

Chapnik, B., G. Desroziers, F. Rabier and O. Talagrand (2003) Properties and first application of an error statistics tuning method in variational assimilation. Submitted for publication in *Q. J. R. Meteorol. Soc.*.

Daley, R. (1992) The Lagged Innovation Covariance : A Performance Diagnostic for Atmospheric Data Assimilation. *Mon. Wea. Rev. 120*, 178-196.

Daley, R. (1993) Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae 11*, 634-647.

Dee, D. P., and A. M. Da Silva (1998) Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc., 124*, 269-295.

Dee, D. P., G. Gaspari, C. Redder, L. Rukhovets and A. M. Da Silva (1999) Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part II: Applications. *Mon. Wea. Rev., 127*, 1835-1849.

Desroziers, G., and S. Ivanov (2001) Diagnosis and adaptive tuning of information error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc. 127*, 1433-1452.

Evensen, G., and P. J. van Leeuwen (2000). An Ensemble Kalman Smoother for Nonlinear Dynamics. *Mon. Wea. Rev. 128*, 1852-1867.

Fisher, M. (2003) *Estimation of Entropy Reduction and Degrees of Freedom for Signal for Large Variational Analysis Systems*, Technical Memorandum No 397, Research Department, Reading, United Kingdom : European Centre for Medium-range Weather Forecasts.

Godbole, S. S. (1974) Kalman Filtering with No A Priori Information About Noise - White Noise Case : Identification of Covariances. *IEEE Trans. Automat. Contr., AC-19*, 561-563.

Hollingsworth, A., and P. Lönnberg (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus 38A*, 111-136.

Hollingsworth, A., and P. Lönnberg (1989) The Verification of Objective Analyses: Diagnostic of Analysis System Performance. *Meteorol. Atmos. Phys, 40*, 3-27.

Houtekamer, P. L., and H. L. Mitchell (2001). A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev. 129*, 123-137.

Kailath, T. (1968) An Innovations Approach to Least-Squares Estimation. Part I: Linear Filtering in Additive White Noise. *IEEE Trans. Automat. Contr, AC-13*, no 6, 646-655.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, *112*, 1177-1194.

Lorenc, A. C. (1997). Quality Control, in Proceedings of Seminar on *Data Assimilation*, September 1996, Reading, United Kingdom : European Centre for Medium-range Weather Forecasts, 251-274.

Mehra, R. K. (1972) Approaches to adaptive filtering, *IEEE Trans. Automat. Contr. 15*, 693-698.

Ménard, R., and L.-P. Chang (2000) Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: χ^2-Validated Results and Analysis of Variance and Correlation Dynamics. *Mon. Wea. Rev. 128*, 2672–2686.

Mitchell, H., C. Charette, C. Chouinard and B. Brasnett (1990) Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Wea. Rev., 118*, 1591-1614.

Rodgers, C. D., 2000, *Inverse Methods for Atmospheric Sounding: Theory and Practice*. London, United Kingdom : World Scientific Publishing Co. Ltd.

Sadiki, W., and C. Fischer (2000) Tuning of a background constraint in a limited 3D-Var System, *CLIVAR Exchanges 5*, no 4, 7-8.

Talagrand, O. (1999) A posteriori evaluation and verification of analysis and assimilation algorithms, Proceedings of Workshop on *Diagnosis of Data Assimilation Systems*, November 1998, Reading, United Kingdom : ECMWF, 17-28.

Talagrand, O. (2003) A posteriori Validation of Assimilation Algorithms, in R. Swinbank, V. Shutyaev and W. A. Lahoz (editors), *Data Assimilation for the Earth System*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 85-95.

Talagrand, O., and F. Bouttier (2000) Internal Diagnostics of Data Assimilation Systems, Proceedings of Seminar on *Diagnosis of Models and Data Assimilation Systems*, September 1999, Reading, United Kingdom: European Centre for Medium-range Weather Forecasts, 407-409.

Wahba, G., D. Johnson, F. Gao and J. Gong (1995) Adaptive tuning of numerical weather prediction models: randomized GCV in three and four dimensional data assimilation. *Mon. Wea. Rev. 123*, 3358-3369.

Weaver, A. T., J. Vialard and D. L. T. Anderson (2003) Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks. *Mon. Wea. Rev. 131*, 1360-1378.