QUARTERLY JOURNAL
OF THE
ROYAL METEOROLOGICAL SOCIETY

Vol. 124 JULY 1998 Part B No. 550

Q. J. R. Meteorol. Soc. (1998), **124**, pp. 1783–1807

The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation

By P. COURTIER[1], E. ANDERSSON[1]*, W. HECKLEY[1], J. PAILLEUX[2], D. VASILJEVIĆ[1], M. HAMRUD[1], A. HOLLINGSWORTH[1], F. RABIER[1] and M. FISHER[1]

[1]*European Centre for Medium-Range Weather Forecasts, UK*
[2]*Centre National de Recherches Météorologiques, France*

(Received 20 December 1996; revised 27 August 1997)

SUMMARY

In the first of this set of three papers, the formulation of the European Centre for Medium-Range Weather Forecasts (ECMWF) implementation of 3D-Var is described. In the second, the specification of the structure function is presented, and the last is devoted to the results of the extensive numerical experimentation programme which was conducted. The 3D-Var formulation uses a spherical-harmonic expansion, much as the ECMWF optimal interpolation (OI) scheme used an expansion of Bessel functions. This formulation is introduced using a convolution algebra over the sphere expressed directly in spectral space. It is shown that all features of the OI statistical model can be implemented within 3D-Var. Furthermore, a non-separable statistical model is described. In the present formulation, geostrophy is accounted for through a Hough-modes separation of the gravity and Rossby components of the analysis increments. As in OI, the tropical analysis remains essentially non-divergent and with a weak mass–wind coupling. The observations used, as well as their specified statistics of errors, are presented, together with some implementation details. In the light of the results, 3D-Var was implemented operationally at the end of January 1996.

KEYWORDS: Data assimilation Numerical weather prediction Objective analysis

1. INTRODUCTION

Following the presentations, at the June 1985 European Centre for Medium-Range Weather Forecasts (ECMWF) workshop on high resolution analysis, of the results obtained by Talagrand and Courtier (1987) and Courtier and Talagrand (1987) with a barotropic vorticity equation, and the shallow-water results of Courtier and Talagrand (1990), a variational data assimilation project was initiated in 1987 in collaboration between the ECMWF and Météo-France (Pailleux 1990). Four-dimensional variational data assimilation (4D-Var) was foreseen as the ultimate goal, with the three-dimensional version (3D-Var) as an important intermediate milestone. Interest in 3D-Var was also stimulated by the difficulties encountered in the late 80's with the use of retrieved temperature and humidity profiles from the TIROS-N Operational Vertical Sounder (TOVS) on board the NOAA satellites (see Andersson *et al.* (1991), Kelly *et al.* (1991) and Flobert *et al.* (1991) for a detailed description of these difficulties). This initiated the 1D-Var project of variational retrieval from the TOVS radiances (Eyre 1989; Eyre and Lorenc 1989; Thépaut and Moll 1990) which was implemented operationally in summer 1992, allowing a positive impact of the TOVS data in the northern hemisphere (Eyre *et al.* 1993). As described by Andersson *et al.* (1994), 1D-Var was conceived as a pre-processing step of 3D-Var or 4D-Var.

* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

Lorenc (1986) showed that Optimal Interpolation (OI) (see Hollingsworth (1987) for a thorough description of OI) and 3D-Var are two algorithms for solving the same linear estimation problem. OI has been in operational use at ECMWF since 1979 (Lorenc 1981; Shaw *et al.* 1987; Undén 1989). The ECMWF 3D-Var formulation has, in many respects, been influenced by the experimental work conducted with OI and, in particular, by the description of the short-range forecast errors by Hollingsworth and Lönnberg (1986) and Lönnberg and Hollingsworth (1986) (referred to as HLLH86 in the following).

Parrish and Derber (1992) and Derber *et al.* (1991) implemented the Spectral Statistical Interpolation analysis scheme (SSI) in June 1991 at the National Centers for Environmental Prediction (NCEP); some subsequent changes were described by Parrish *et al.* (1995). SSI and 3D-Var are very similar variational algorithms. They nevertheless differ in several respects which will be mentioned during the course of this paper. The Meteorological Office (Lorenc 1995) and the Atmospheric Environment Service (Gauthier *et al.* 1996) are also developing 3D-Var schemes, the latter being very close to SSI or the ECMWF 3D-Var while the former originally did not make use of the spherical harmonics for representing the statistics of short-range forecast errors (Lorenc 1992). The Data Assimilation Office has devised the PSAS algorithm (Cohn *et al.* 1998) which is in duality with 3D-Var (Courtier 1997).

The present paper describes the formulation of the 3D-Var algorithm. Two companion papers, Rabier *et al.* (1998) and Andersson *et al.* (1998) (referred to subsequently as Part II and Part III), describe the calculation of the background-error statistics, and the experimental results. The extensive experimental programme was followed by operational implementation on 30 January 1996. After this introduction, the second section summarizes the 3D-Var principle. The third section presents the background-term formulation. The fourth section describes the observations used and the specification of their error statistics. The fifth section summarizes some implementation features. Numerical results in idealized cases are followed by a brief conclusion restricted to the formulation aspects. The notation used in the paper follows that used by Ide *et al.* (1997).

2. 3D-VAR

In its incremental formulation (Courtier *et al.* 1994), 3D-Var attempts to minimize the objective function J

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x}}_{J_{\mathrm{b}}} + \underbrace{\frac{1}{2}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})}_{J_{\mathrm{o}}}, \tag{1}$$

where $\delta\mathbf{x}$ is the increment and, at the minimum, the resulting analysis increment $\delta\mathbf{x}^{\mathrm{a}}$ is added to the background $\mathbf{x}^{\mathrm{b}}$ in order to provide the analysis $\mathbf{x}^{\mathrm{a}}$.

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}^{\mathrm{a}}. \tag{2}$$

$\mathbf{B}$ is the covariance matrix of background error while $\mathbf{d}$ is the innovation vector

$$\mathbf{d} = \mathbf{y}^{\mathrm{o}} - H\mathbf{x}^{\mathrm{b}}, \tag{3}$$

where $\mathbf{y}^{\mathrm{o}}$ is the observation vector. $\mathbf{H}$ is a suitable linear approximation of the observation operator H in the vicinity of $\mathbf{x}^{\mathrm{b}}$, and $\mathbf{R}$ is the covariance matrix of observation errors.

The incremental formulation of 3D-Var consists therefore of solving for $\delta\mathbf{x}$ the inverse problem defined by the (direct) observation operator $\mathbf{H}$, given the innovation vector $\mathbf{d}$. The gradient of J is obtained by differentiating Eq. (1) with respect to $\delta\mathbf{x}$

$$\nabla J = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})\delta\mathbf{x} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}\,. \tag{4}$$

At the minimum, the gradient of the objective function vanishes, thus from Eq. (4) we obtain the classical result that minimizing the objective function defined by Eq. (1) is a way of solving (for $\delta\mathbf{x}^{\mathrm{a}}$) the linear system

$$(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})\delta\mathbf{x}^{\mathrm{a}} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}\,. \tag{5}$$

The solution of 3D-Var also satisfies the OI equation (see, for example, Lorenc (1986) for this standard result)

$$\delta\mathbf{x}^{\mathrm{a}} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{d}\,. \tag{6}$$

$\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ may therefore be interpreted as the square matrix of the covariances of background errors in observation space while $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ is the rectangular matrix of the covariances between the background errors in model space and the background errors in observation space.

Most (if not all) implementations of OI rely on a statistical model for describing $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ (see HLLH86 and Bartello and Mitchell (1992)). 3D-Var uses the observation operator $\mathbf{H}$ explicitly and if, as OI, it requires a statistical model, this is only for describing the statistics of background errors in model space. Consequently, in 3D-Var it turns out to be easier from an algorithmic point of view to make use of observations, such as TOVS radiances, which depend nonlinearly on the basic analysis-variables.

3. Formulation of the background term

In principle, any covariance matrix $\mathbf{B}$ (i.e. positive definite) may be specified in Eq. (1). In operational practice, $\delta\mathbf{x}$ being a vector of size 10^5 to 10^7, $\mathbf{B}$ is a matrix which contains 10^{10} to 10^{14} elements. As a consequence, it is not tractable for dealing with any positive definite matrix. Furthermore, as noted by Dee (1995), the available statistical information necessary for specifying the elements of $\mathbf{B}$ is severely limited. All the information useful for the specification of the statistics of a Kalman filter has to be extracted from the sequence of innovation vectors. This corresponds, in operational meteorology, to 10^7 to 10^8 observations per annum. Even relying on some form of ergodicity*, the number of elements of the $\mathbf{B}$ matrix which could be estimated stably is several orders of magnitude less. A statistical model has therefore to be defined. In the rest of this section we shall build such a model which incorporates the features observed by HLLH86. The reader is referred to Part II (Rabier *et al.* 1998) for the actual specification of the statistics.

The J_{b} formulation is now introduced, using tensorial analysis since it naturally extends to the infinite dimension case. This considerably simplifies the notation and allows interpretation with correlation functions. It is possible to present most of the following with matrix notation, see Heckley *et al.* (1993).

(*a*) *Univariate case, bidimensional*

$\delta\mathbf{x}$ is considered in this subsection as a scalar field defined over the sphere Σ. $\delta\mathbf{x}$ may be decomposed into spherical harmonics Y_n^m

$$\delta\mathbf{x} = \sum_n \sum_{m=-n}^{n} \delta\mathbf{x}_n^m Y_n^m\,. \tag{7}$$

* A stochastic process is said to be ergodic if it has the property that sample (or time) averages formed from an observed record of the process may be used as an approximation for the corresponding ensemble averages.

$\delta\mathbf{x}$ is a real field, thus

$$\delta\mathbf{x}_n^m = \delta\mathbf{x}_n^{-m*} = \delta\mathbf{x}_{n_\mathrm{r}}^m + i\delta\mathbf{x}_{n_\mathrm{i}}^m,$$

where the asterisk denotes the complex conjugate, the subscripts i and r indicate respectively the imaginary and real part, and $i^2 = -1$. The normalization for the spherical harmonics is the usual one in meteorology

$$\frac{1}{4\pi}\int_\Sigma Y_n^m Y_n^{m*} \mathrm{d}\Sigma = 1\,. \tag{8}$$

Let us consider the random field ξ which we assume unbiased for the sake of simplicity $\langle\xi\rangle = 0$, $\langle.\rangle$ standing for the mathematical expectation. We introduce the covariance tensor $\mathbf{T} = \langle\xi\xi^{\mathrm{T}}\rangle$. A covariance tensor has two entries, and the covariance of two fields x and y is denoted by $\mathbf{T}(x, y)$. If P and Q are two points of the sphere Σ, the covariance between these two points is defined as being $\mathbf{T}(\delta_P, \delta_Q)$, where δ_P and δ_Q are the Dirac distributions at points P and Q respectively. $\mathbf{T}(\delta_P, \delta_P)$ is the variance of the random field ξ at point P. In this section, we assume that the mathematical objects we manipulate are properly defined and finite in the appropriate functional and probabilistic spaces. As an example, the variance of any random field ξ is not necessarily well defined, and its existence has some implications for the underlying probability law. This difficulty was discussed by Gaspari and Cohn (1999) and the authors of references therein.

(i) *Isotropic covariances.* The fundamental result at the heart of 3D-Var (see for example Boer (1983)), is that if $\mathbf{T}$ is isotropic, i.e. invariant by all rotations over the sphere, then the spherical harmonics are orthogonal for the metric defined by $\mathbf{T}$, or in other words, the two spherical-harmonic fields Y_n^m and $Y_{n'}^{m'*}$ are correlated only if they are equal

$$\mathbf{T}(Y_n^m, Y_{n'}^{m'*}) = \delta_{n-n'}\delta_{m-m'}b_n\,, \tag{9}$$

the modal variance b_n being independent of the zonal wave-number m. $\delta_{n-n'}$ is the Kronecker symbol, equal to 1 if $n = n'$, and 0 otherwise. The result is an immediate consequence of the fact that spherical harmonics provide an irreducible representation of the rotation group of the sphere. Boer (1983) presented an elementary proof which relied on the addition theorem of zonal spherical harmonics. This proof, sketched below, is by construction and provides an interpretation of the b_n. As $\mathbf{T}$ is isotropic, the covariance between two points P and Q of the sphere is a function of their distance over the sphere only, and therefore of their angular separation θ.

$$\mathbf{T}(\delta_P, \delta_Q) = f(\theta)\,; \tag{10}$$

f may be decomposed in terms of the zonal spherical harmonics $Y_n^0 = P_n^0$:

$$f(\theta) = \sum_n f_n P_n^0(\cos\theta)\,. \tag{11}$$

The Legendre polynomials P_n^0 are normalized and, from Eq. (8), one immediately finds

$$\frac{1}{2}\int_0^\pi P_n^0(\cos\theta)^2 d\cos\theta = 1\,. \tag{12}$$

This is why we keep the superscript 0, to distinguish between the normalization commonly used in mathematics, $P_n(1) = 1$, and the normalization used here which implies $P_n^0(1) = \sqrt{2n+1}$.

The addition theorem for zonal spherical harmonics relates b_n to f_n (Boer 1983; Courtier 1987; Wunsch and Stammer 1995)

$$f_n = b_n\sqrt{2n+1}\,. \tag{13}$$

The variance at any point P is then

$$\begin{aligned}\mathbf{T}(\delta_P, \delta_P) = f(0) = \sum_n f_n P_n^0(1) &= \sum_n f_n\sqrt{2n+1} \\ &= \sum_n b_n(2n+1)\,. \end{aligned} \tag{14}$$

Thus b_n is the modal variance while $(2n+1)b_n$ is the total variance for a given total wave-number n, and thus provides the power spectrum.

The metric defined by Eq. (8) allows identification of $\mathbf{T}(\bullet, \delta\mathbf{x})$ with a field which may be interpreted as the convolution of the field $\delta\mathbf{x}$ with the isotropic field centred at the northern pole, and therefore a zonally uniform field, which has the value $f(\theta)$ for a given co-latitude θ (see Yaglom (1987)). Denoting the convolution by an asterisk, the addition theorem for the zonal spherical harmonic provides the result

$$(f * \delta\mathbf{x})_n^m = (\mathbf{T}(\bullet, \delta\mathbf{x}))_n^m = \frac{1}{\sqrt{2n+1}}\delta\mathbf{x}_n^m f_n\,. \tag{15}$$

f is said to *represent* the isotropic covariance tensor $\mathbf{T}$ (Gaspari and Cohn 1999). Since the f_n may be interpreted as a variance, they are necessarily positive. Conversely, any function f with positive Legendre expansion coefficients is the representation of a covariance tensor. This is the Böchner–Schwartz theorem (Gel'Fand and Vilenkin 1964, p. 157) in the particular case of the sphere. Let us consider another isotropic tensor $\mathbf{U}$, represented by a function g. Equation (15) applied twice allows definition of the tensor $\mathbf{U}\times\mathbf{T} = \mathbf{T}\times\mathbf{U}$ which is represented by the function $f * g$ with

$$(f * g)_n = \frac{1}{\sqrt{2n+1}} f_n g_n\,. \tag{16}$$

$(\mathbf{U}\times\mathbf{T})(\bullet, \delta\mathbf{x})$ is identified with the field which may be obtained as the convolution of the field $\delta\mathbf{x}$, first with the zonally uniform field which, for a given co-latitude θ, has $f(\theta)$ for its value, and then with the zonally uniform field which has $g(\theta)$ for its value.

The Dirac distribution at the northern pole δ_N has the following spherical-harmonic expansion

$$\begin{aligned}\delta_{N_n}^m = \frac{1}{4\pi}\int_\Sigma \delta_{N_n}^m Y_n^{m*} d\Sigma &= \delta_{m-0} P_n^0(0) \\ &= \delta_{m-0}\sqrt{2n+1}\,. \end{aligned} \tag{17}$$

From Eqs. (16) and (17), one finds the classical result that the Dirac distribution at the northern pole δ_N is the neutral element of the convolution. Define the inverse of the isotropic covariance tensor $\mathbf{T}$ represented by f as the isotropic tensor denoted by $\mathbf{T}^{-1}$ and represented by a function g such that $f * g = \delta_N$. Combining Eqs. (16) and (17), one finds

$$g_n = \frac{2n+1}{f_n}\,, \tag{18}$$

and then

$$(\mathbf{T}^{-1}(\bullet, \delta\mathbf{x}))_n^m = (g * \delta\mathbf{x})_n^m = \frac{\sqrt{2n+1}}{f_n}\delta\mathbf{x}_n^m . \tag{19}$$

The expression of J_b immediately follows

$$\begin{aligned} J_b(\delta\mathbf{x}) = \frac{1}{2}\mathbf{T}^{-1}(\delta\mathbf{x}, \delta\mathbf{x}) &= \frac{1}{2}\frac{1}{4\pi}\int_\Sigma \delta\mathbf{x}\mathbf{T}^{-1}(\bullet, \delta\mathbf{x})\mathrm{d}\Sigma \\ &= \frac{1}{2}\sum_n \sum_{m=-n}^{n} \delta\mathbf{x}_n^m \delta\mathbf{x}_n^{m*}\frac{\sqrt{2n+1}}{f_n} \\ &= \frac{1}{2}\sum_n \sum_{m=-n}^{n} \delta\mathbf{x}_n^m \delta\mathbf{x}_n^{m*}\frac{1}{b_n} \\ &= \frac{1}{2}\sum_n \frac{1}{b_n}\left\{\delta\mathbf{x}_n^{0^2} + 2\sum_{m=1}^{n}\left(\delta\mathbf{x}_{n_r}^{m^2} + \delta\mathbf{x}_{n_i}^{m^2}\right)\right\} . \end{aligned} \tag{20}$$

Following Daley (1991), the length scale L_s of the isotropic covariance tensor may be defined as

$$L_s^2 = -2\frac{f(0)}{\Delta f(0)} = 2a^2\frac{\sum_n f_n\sqrt{2n+1}}{\sum_n f_n n(n+1)\sqrt{2n+1}} , \tag{21}$$

Δ being the Laplacian operator and a the radius of the sphere Σ.

(ii) *Preconditioning.* In practice, it is necessary to precondition the minimization problem in order to obtain a quick convergence. As the Hessian of the objective function is not accessible, Lorenc (1988) suggested the use of the Hessian of the background term J_b. In practice, such a preconditioning may be implemented either by a change of metric in the space of the control variable, or by a change of control variable. As the minimization algorithms have generally to evaluate several inner products, it was found more efficient to implement a change of variable. Algebraically, this requires the introduction of a variable $\mathbf{z}$ such that

$$J_b = \frac{1}{2}\mathbf{z}^T\mathbf{z}. \tag{22}$$

Comparing Eqs. (1) and (22) shows that $\mathbf{z} = \mathbf{B}^{-\frac{1}{2}}\delta\mathbf{x}$ would satisfy the requirement. This procedure may be applied directly to Eq. (20). Nevertheless, in the following, we shall introduce the square root of an isotropic covariance tensor, denoted by $\mathbf{T}^{\frac{1}{2}}$, since this provides an interpretation which will be of some use later. We aim to find the representation h of $\mathbf{T}^{\frac{1}{2}}$ such that $h * h = f$.

$$(h * h)_n = \frac{1}{\sqrt{2n+1}}h_n^2 \tag{23}$$

which implies

$$(f^{\frac{1}{2}})_n = \quad h_n \quad = \sqrt{\sqrt{2n+1}\,f_n} \tag{24}$$

$$(f^{-\frac{1}{2}})_n = (h^{-1})_n = \frac{(2n+1)^{\frac{3}{4}}}{\sqrt{f_n}} . \tag{25}$$

Introducing $\mathbf{z} = \mathbf{T}^{-\frac{1}{2}}(\bullet, \delta\mathbf{x})$, one finds the expression for J_b

$$\begin{aligned} J_b &= \frac{1}{2}\frac{1}{4\pi}\int_\Sigma \delta\mathbf{x}\mathbf{T}^{-1}(\bullet, \delta\mathbf{x})\mathrm{d}\Sigma \\ &= \frac{1}{2}\sum_n \sum_{m=-n}^{n} \delta\mathbf{x}_n^m \delta\mathbf{x}_n^{m*} \frac{\sqrt{2n+1}}{f_n} \\ &= \frac{1}{2}\sum_n \sum_{m=-n}^{n} \left(\frac{1}{\sqrt{2n+1}}\frac{(2n+1)^{\frac{3}{4}}}{\sqrt{f_n}}\delta\mathbf{x}_n^m\right)\left(\frac{1}{\sqrt{2n+1}}\frac{(2n+1)^{\frac{3}{4}}}{\sqrt{f_n}}\delta\mathbf{x}_n^m\right)^* \\ &= \frac{1}{2}\frac{1}{4\pi}\int_\Sigma \mathbf{T}^{-\frac{1}{2}}(\bullet, \delta\mathbf{x})\mathbf{T}^{-\frac{1}{2}}(\bullet, \delta\mathbf{x})\mathrm{d}\Sigma \\ &= \frac{1}{2}\frac{1}{4\pi}\int_\Sigma \mathbf{z}^2\mathrm{d}\Sigma\,. \end{aligned} \tag{26}$$

Lorenc (1995) describes a 3D-Var system which makes use of recursive filters to evaluate the convolution of $\mathbf{z}$ by $\mathbf{T}^{\frac{1}{2}}$ in grid-point space, which is the only convolution necessary to recover $\delta\mathbf{x}$.

(iii) *Discrete case.* In practice, the fields have to be represented with a finite spherical-harmonics expansion. At ECMWF, a triangular truncation ($n \leqslant n_{\max}$) is used, but the above development would also be valid for other truncation types (rhomboidal or trapezoidal). All the above computations are currently made in Laplace space, the summations over n being limited to $n_{\max}$.

For a field $\mathbf{z}$ decomposed in a finite spherical-harmonic expansion, one has

$$\frac{1}{4\pi}\int_\Sigma \mathbf{z}^2\mathrm{d}\Sigma = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} \mathbf{z}_n^m \mathbf{z}_n^{m*} = \sum_{n=0}^{n_{\max}} \left\{ \mathbf{z}_n^{0^2} + 2\sum_{m=1}^{n} \left(\mathbf{z}_{n_r}^{m^2} + \mathbf{z}_{n_i}^{m^2}\right)\right\}\,. \tag{27}$$

In spectral models, a collocation grid is used for evaluating the nonlinear terms of the equations of evolution (Bourke 1972). It is defined as a set of $i_{\max}$ co-latitudes θ_i used for the Gaussian quadrature, with Gaussian weights ω_i in the north–south direction and a set of $j_{\max}$ longitudes λ_j used for the Fourier transform in the east–west direction. The Parseval formula has a discrete equivalent, provided the grid has enough resolution ($j_{\max} \geqslant 1 + 2n_{\max}$ and $i_{\max} \geqslant 1 + n_{\max}$).

$$\frac{1}{4\pi}\int_\Sigma \mathbf{z}^2\mathrm{d}\Sigma = \sum_{n=0}^{n_{\max}} \left\{ \mathbf{z}_n^{0^2} + 2\sum_{m=1}^{n} \left(\mathbf{z}_{n_r}^{m^2} + \mathbf{z}_{n_i}^{m^2}\right)\right\} = \frac{1}{j_{\max}} \sum_{i=1}^{i_{\max}} \omega_i \sum_{j=1}^{j_{\max}} \mathbf{z}(\lambda_j, \theta_i)^2\,. \tag{28}$$

Courtier and Naughton (1994) showed that the Parseval formula remains valid to machine precision with an appropriately defined reduced Gaussian grid. Therefore it would also have been possible to evaluate the summation in grid-point space which, as we shall see in subsection (vi), introduces further flexibility into the formulation.

(iv) *Isotropic correlations.* Rutherford (1972) noted that, in operational meteorology, the assumption of isotropic correlations is by far superior to isotropic covariances. It allows account to be taken of the contrast between data-rich and data-poor areas which leads to different quality in the subsequent short-range forecasts. Such a feature was implemented in the operational ECMWF OI (Shaw *et al.* 1987).

The random field ξ is, as above, assumed unbiased, but we now assume that it is the correlation which is isotropic, rather than the covariances. Let us denote the variances of

ξ by σ_b^2. The normalized field $\xi' = \xi/\sigma_b$ is of isotropic covariances with variance equal to 1. It is therefore possible to apply the result of the previous paragraph and

$$J_b(\delta\mathbf{x}) = \sum_n \sum_{m=-n}^{n} \delta\mathbf{x}'^m_n \delta\mathbf{x}'^{m*}_n \frac{\sqrt{2n+1}}{f_n}, \tag{29}$$

with

$$\delta\mathbf{x}' = \frac{\delta\mathbf{x}}{\sigma_b}.$$

f is the representation of an isotropic tensor of variance equal to 1. Equation (14) relates the variance of an isotropic covariance tensor to the f_n, therefore

$$\sum_n f_n \sqrt{2n+1} = 1. \tag{30}$$

This feature is implemented in the ECMWF 3D-Var but not in the NCEP SSI in which the isotropy of the covariances is relaxed in spectral space (see subsection (viii) below).

(v) *Compactly supported correlations.* In most statistical models used so far in OI, the horizontal correlations essentially vanish beyond a distance of typically 2000 km in the troposphere (see e.g. HLLH86). HLLH86 used a decomposition in Bessel-function series with a radius of the underlying cylinder of 3000 km while Bartello and Mitchell (1992) used a 3800 km radius. This may also be a desirable feature of the statistical model used in 3D-Var. Gaspari and Cohn (1998) developed the theory for the construction of compactly supported correlations in two and three dimensions that we follow in the remainder of this section. The basic idea is to define a compactly supported function and through self-convolution to obtain a correlation function which, by construction, is compactly supported.

Given a covariance tensor $\mathbf{T}$ represented by the function f, we have seen with Eq. (24) how to generate the function $f^{\frac{1}{2}}$ which represents the covariance tensor $\mathbf{T}^{\frac{1}{2}}$. Define the function $\widetilde{f}^{\frac{1}{2}}$, resetting the function $f^{\frac{1}{2}}$ to 0 beyond a given distance. $\widetilde{f} = \widetilde{f}^{\frac{1}{2}} * \widetilde{f}^{\frac{1}{2}}$ is a correlation function since its Legendre coefficients expansion are positive by construction, as seen from Eq. (16). Furthermore, $\widetilde{f}$ is compactly supported. However, the Legendre expansion cannot be finite (Gaspari and Cohn 1998). Nevertheless, we found in practice that finite expansion was obtained with $\widetilde{f}$ compactly supported within machine precision. This is further discussed in Part II of this paper (Rabier *et al.* 1998).

This feature was implemented neither at ECMWF nor at NCEP, and we shall see in Part II that it is most likely a weakness of the original implementation. It has, subsequently, been implemented in the 1997 revision of 3D-Var at ECMWF.

(vi) *Horizontally variable length-scale.* Generally, the operational implementation of OI used different length-scales depending on the geographic location, thus allowing for a broader length-scale in data-poor areas than in data-rich areas. Empirically, this was found useful in operational practice and justified the use of a 2-dimensional Kalman filter by Bouttier (1994).

Consider two isotropic covariance tensors $\mathbf{T}_1$ and $\mathbf{T}_2$ and two complementary subdomains Σ_1 and Σ_2 of the sphere Σ. Define $\mathbf{z}_1 = \mathbf{T}_1^{-\frac{1}{2}}(\bullet, \delta\mathbf{x})$ and $\mathbf{z}_2 = \mathbf{T}_2^{-\frac{1}{2}}(\bullet, \delta\mathbf{x})$. J_b may now be defined as

$$J_b = \frac{1}{4\pi}\int_{\Sigma_1} \mathbf{z}_1^2 d\Sigma + \frac{1}{4\pi}\int_{\Sigma_2} \mathbf{z}_2^2 d\Sigma. \tag{31}$$

In the discrete case, the two integrals may be evaluated in grid-point space, making use of the Parseval formula. This could easily be generalized to a few subdomains, the limitation being the induced cost, since a set of spectral transforms is required per subdomain. Neither the ECMWF 3D-Var nor the NCEP SSI have incorporated this feature so far. Nevertheless, we shall see in Part III (Andersson *et al.* 1998) that it would be desirable, the typical length-scale of the short-range-forecast errors being shorter over data-dense areas like the northern hemisphere continents than over data-poor areas like the oceans (Bouttier 1994).

(vii) *Relationship with Bessel-function expansions.* As OI was implemented solving Eq. (6) locally, the statistical model only needed to be defined on the tangent plane to the sphere. Then, following Gandin (1963) and Rutherford (1972), HLLH86 and Bartello and Mitchell (1992) used Bessel-function expansions to represent the correlation functions of the OI statistical model. In order to be able to relate the 3D-Var statistical model to those used with OI, it is necessary, therefore, to study its behaviour in the limit of small angular separation, when the tangent-plane approximation makes sense, i.e. when the horizontal distance is small compared to the radius of the sphere.

From a formula (9.1.71, p. 362) of Abramowitz and Stegun (1965) or formula A.2.2 of Courtier and Naughton (1994), we find that for large n and small θ, such that the product $n\theta$ remains fixed,

$$P_n^0(\cos\theta) \approx \sqrt{2n+1}\,\mathcal{J}_0(n\theta)\,. \tag{32}$$

Equation (11) then becomes

$$f(\theta) \approx \sum_n f_n \sqrt{2n+1}\,\mathcal{J}_0(n\theta)\,. \tag{33}$$

HLLH86 required the derivative of the correlation to be zero at some distance d (3000 km in their case; 3800 km in the work of Bartello and Mitchell (1992)). This allows Eq. (33) to be rewritten as

$$f(\theta) \approx \sum_n f_n \sqrt{2n+1}\,\mathcal{J}_0\left(\frac{nd}{a}\frac{a\theta}{d}\right) \tag{34}$$

and therefore $k_n = nd/a$ to be interpreted as the wave number in the Bessel-function expansion since $-(k_n/d)^2 = -(n/a)^2$ is the eigenvalue of the Laplacian operator in polar coordinates with $\mathcal{J}_0(n\theta)$ as eigenfunction. In Part II the power spectrum specified in 3D-Var will be compared with the results of previous studies, using Eq. (34).

HLLH86 relate k_n to an equivalent total wave-number n using

$$(k_n/d)^2 = n(n+1)/a^2\,.$$

In practice, the difference is half a wave number, except for wave-number zero where there is no difference and where the interpretation does not make a lot of sense.

(viii) *Relaxing the isotropy.* In Eq. (20), which defines J_b in spectral space for an isotropic covariance tensor, it is possible to introduce a dependance of the b coefficients not only with total wave-number n, but with both total wave-number n and zonal wave-number m, thus still keeping the spherical harmonics orthogonal for the metric defined by the covariance tensor. This is a feature operationally implemented in the NCEP SSI.

Let us assume that the tensor $\mathbf{T}$ is invariant to rotation about the polar axis. As a consequence, the different zonal wave-numbers m are decoupled from each other, and in spectral space, the matrix of $\mathbf{T}$ is block diagonal, with a block for each wave number m. If in addition, $\mathbf{T}$ is symmetric with respect to a meridian plane, the imaginary and real parts of a given wave-number m are decoupled, each block being equal. Furthermore, the

symmetry with respect to the equator decouples the odd from the even harmonics (the parity of a harmonic is defined as the parity of $n - m$, an even harmonic is symmetric with respect to the equator while an odd one is antisymmetric).

In the tangent plane to the sphere at given co-latitude θ, the covariance tensor may be approximated to second order in both coordinate directions. The iso-covariances are then ellipses. The above symmetry-properties require the two axes of the ellipse to be parallel to the axes of coordinates. In addition, at the northern pole, the ellipse reduces to a circle.

The orthogonality of the spherical harmonics for the metric defined by the covariance tensor implies some further properties that have not been investigated. For example, the variances may also have a latitudinal variability, since the spherical harmonic expansion of the Dirac distribution at the northern pole depends only on the zonal harmonics whereas the Dirac distribution at the equator depends on all zonal wave-numbers.

(*b*) *Univariate case, tridimensional*

As does the ECMWF model, 3D-Var exploits the shallowness of the atmosphere by assuming the metric defined over the space of atmospheric fields to be independent of the vertical (Müller 1989). The equipotentials are assumed spherical and therefore at the same distance a from the centre of the earth (Phillips 1973). A vertical coordinate is then chosen which, say, varies from zero at the top of the atmosphere to one at the bottom. This vertical coordinate may be the σ-coordinate as at NCEP or the hybrid η-coordinate of Simmons and Burridge (1981).

It is important to realize at this stage that the notion of horizontal isotropy is not intrinsic and depends on the choice of vertical coordinate. This is a consequence of the degeneracy of the metric in the vertical direction which implies that for any vertical coordinate, the coordinate surfaces are horizontal, i.e. orthogonal to the normal to the equipotentials. If there is no ambiguity, in what follows we shall use 'horizontal isotropy' with respect to the η-surfaces.

A three-dimensional atmospheric field x may be expanded in terms of the spherical harmonics to obtain

$$\mathbf{x} = \sum_n \sum_{m=-n}^{n} \mathbf{x}_n^m(\eta) Y_n^m \,. \tag{35}$$

The covariance tensor $\mathbf{T}$ is assumed invariant to rotation. Fixing the value of the vertical coordinate η in the first entry of $\mathbf{T}$ and η' in the second entry, one obtains a covariance tensor defined over the sphere which, being the trace of an isotropic covariance tensor, is itself isotropic. Therefore one may apply the bidimensional result of Eq. (9)

$$\mathbf{T}(Y_n^m \delta_\eta, Y_{n'}^{m'*} \delta_{\eta'}) = \delta_{n-n'} \delta_{m-m'} b_n(\eta, \eta') \,. \tag{36}$$

$b_n(\eta, \eta')$ can be interpreted as the covariance between level η and level η' for the horizontal scale associated with the total wave-number n. In the vertically discretized case, the isotropic covariance tensor has a block-diagonal expression. A vertical covariance matrix $\mathbf{B}_n$ consisting of the $b_n(\eta, \eta')$ may be defined for every total wave-number n. All the bidimensional results of the previous section may then be generalized to the three-dimensional case. As an illustration, the vertical covariance matrix, denoted by $\mathbf{B}$, is obtained from the matrices $\mathbf{B}_n$ with

$$\mathbf{B} = \sum_n \mathbf{B}_n (2n + 1) \,, \tag{37}$$

which is the equivalent of Eq. (14).

Remark 1. Such a statistical model is, in general, non-separable in that the covariance between two points cannot be expressed as the product of a function of the horizontal distance with a function of the vertical separation. Separability implies that the matrices $\mathbf{B}_n$ are obtained as the product of a vertical covariance matrix, independent of n, with a spectrum which is independent of the vertical. The ECMWF OI relied on this hypothesis, apart from a large-scale term in the geopotential covariance which was uncorrelated with streamfunction or velocity potential.

Remark 2. A hierarchy of non-separable models may be introduced, ranging between the general formulation and the separable formulation. If, in the separable formulation, one introduces a spectrum which depends on the vertical, the length scale could vary with the vertical. Another possibility would be to assume the EOF of the matrices $\mathbf{B}_n$ to be independent of n. None of these formulations were found to be in close agreement with the empirical statistics which is why we retained the general formulation.

Remark 3. As the ECMWF OI relied on a Bessel-function expansion, it would have been straightforward (but of significant cost) to implement a non-separable statistical model similar to that of 3D-Var.

Remark 4. The isotropic correlations are obtained, as in the bidimensional case, by first dividing in grid-point space by the standard deviations of background errors. This procedure may be generalized by dividing in grid-point space by the (spatially variable) square root of the vertical covariance matrix of background error. This introduces spatial variability of the vertical covariance matrix. This generalization is not in the first implementation of the ECMWF 3D-Var.

(*c*) *Winds*

(i) *Bidimensional.* The wind field is uniquely decomposed over the sphere into divergent and rotational components obtained from the velocity potential χ and streamfunction ψ respectively. Isotropy is assumed for the covariance tensor of the doublet (χ, ψ) which implies a block-diagonal structure in spectral space with a 2–2 square matrix which depends on the the total wave-number n only for every doublet $(\chi^m_{nr}, \psi^m_{nr})$ or $(\chi^m_{ni}, \psi^m_{ni})$. No correlation between the rotational and divergent component then results in diagonal 2–2 matrices.

Isotropic correlations are considered as in the scalar case by dividing in grid-point space by the standard deviations of background errors. However, it is not equivalent to perform this normalization for the wind field as compared to the streamfunction and velocity potential. In order to keep the velocity potential and the streamfunction independent, the normalization should not be performed on the wind field. In practice, this is critical since the divergence field is an order of magnitude smaller than the vorticity field.

(ii) *Tridimensional.* The above results are generalized in practice to the tridimensional case following what was done for the scalar fields. Nevertheless, there is a hidden difficulty which we detail now. The wind field is defined as the velocity of an air parcel along the equipotentials. However, in 3D-Var, as in the ECMWF model, the horizontal derivatives are taken along the coordinate surfaces. Consequently, the streamfunction, the velocity potential, as well as the vorticity and the divergence are concepts which depend on the choice of vertical coordinate. No practical problems have been found so far. However, they are likely to arise at high resolution close to orography in relation to the multivariate formulation.

(*d*) *Multivariate formulation*

The multivariate formulation of 3D-Var has been discussed in detail by Courtier *et al.* (1993) and in this subsection we shall summarize the main features. Parrish and Derber (1992) diagnosed the balanced part of the mass field from the vorticity field using a variant of the balance equation accounting for some divergence effects. The control variable consisted then of the vorticity, the divergence and the unbalanced part of the mass field. We have followed another route, remaining closer to nonlinear normal-mode initialization, and this is described below. The results of the experimentation presented in part III revealed several weaknesses, leading to a revised formulation described by Bouttier *et al.* (1997) and which is very close to that of Parrish and Derber (1992) for the multivariate aspects.

In the vicinity of a state of rest, the slow manifold is tangent to the Rossby linear manifold R (see, for example, Daley (1991) for a thorough description of the concepts used here and Courtier *et al.* (1993) for the implications for 3D-Var). As 3D-Var uses a spectral representation of the fields, the Rossby linear manifold is readily accessible and the projection on R consists of linear normal-mode initialization. $\delta\mathbf{x}$ is separated into three components, namely a Rossby component $\delta\mathbf{x}_R$, a gravity component $\delta\mathbf{x}_g$, and a univariate component $\delta\mathbf{x}_u$.

In mid-latitudes, the univariate component is set to zero, while the gravity component is penalized to a degree which depends both on the total wave-number n and the vertical scale (the index being the rank of the eigenvector of the vertical covariance matrix). This dependence on both horizontal and vertical scale has been found in the empirical statistics and, for example, the dependence on horizontal scale was implemented in the ECMWF OI following the results of HLLH86.

In the tropics, the gravity and Rossby components are set to zero. The analysis is then univariate, but the divergence field of the univariate component remains penalized. In the latitude band 15°–30° a smooth transition from the mid-latitude to the tropical analysis is implemented. This mimics the ECMWF OI.

The vertical staggering of the ECMWF model introduces a spurious $2\delta\eta$ noise in the vertical while going back from the mass variable $P = \phi + RT\ln(p_s)$ to temperature T and surface pressure p_s, ϕ being the geopotential. The noise reached an amplitude of a few degrees in the stratosphere. This was solved by the use of a pseudo-inverse of the integration of the hydrostatics consistent with the vertical covariance matrices (see Courtier *et al.* (1993) for a comprehensive discussion).

In order to avoid double counting in the evaluation of J_b, only the contribution of vorticity accounts for the Rossby component, only the contribution of divergence and mass accounts for the gravity component while the three are considered for the univariate component. The first two total wave-numbers of the mass field are also considered in the evaluation of J_b of the Rossby component since they depend weakly, if at all, on vorticity.

The ECMWF implementation of OI required a special treatment of the tides (Wergen 1988); similar difficulties have been encountered in the early experimentation of 3D-Var. The 12-hour periodicity of the radiosonde network, the two helio-synchronous NOAA satellites in quadrature and the six-hour cycle of the assimilation are all in potential resonance with the natural frequencies of the atmospheric tides. Diurnal and semi-diurnal oscillations of the assimilation increments were clearly visible in a harmonic analysis performed over two weeks of assimilations. In essence, the modification introduced to the background term is an adequate penalization of the tidal contribution performed in spectral space, followed by nonlinear normal mode initialization of the increments (see section (*a*)).

Surface-pressure forecasts plotted every time-step at selected grid-points showed that the above formulation was able to control fast surface-pressure oscillations to a large extent with a maximum amplitude in the remaining oscillations of a fraction of a millibar. The introduction in the objective function of a term J_c measuring the distance to the slow manifold as in Courtier and Talagrand (1990) eliminated the remaining oscillations, even close to orography. This term comprises the square norm of the tendencies of the gravity waves.

4. Observations

The observation operators provide the link between the analysis variables and the observations (Lorenc 1986; Pailleux 1990). The operator H in Eq. (3) signifies the ensemble of operators transforming the control variable $\mathbf{z}$ into the equivalent of each observed quantity, $\mathbf{y}^o$, at observation locations. The 3D-Var implementation allows H to be (weakly) nonlinear, which is seen to be an advantage for the use of TOVS radiance data, for example.

In this section, we outline the content of the observation operators and describe the observational data used in 3D-Var.

(*a*) *Observation operators*

The operator H is subdivided into a sequence of operators, each one of which performs part of the transformation from control variable to observed quantity: (i) the inverse change of variable converts from control variables to model variables; (ii) the inverse spectral transforms put the model variables on the model's reduced Gaussian grid; (iii) a 12-point bi-cubic horizontal interpolation gives vertical profiles of model variables at observation points. The surface fields are interpolated bi-linearly to avoid spurious maxima. Steps (i) to (iii) are in common for all data types. Thereafter follows (iv) vertical integration of, for example, the hydrostatic equation to form geopotential, and of the radiative transfer equation to form radiances (if applicable), and (v) vertical interpolation to the level of the observations. The vertical operations depend on the variable. The vertical interpolation is linear in pressure for temperature and specific humidity, and linear in the logarithm of pressure for wind. The vertical interpolation of geopotential is similar to that for wind (in order to preserve geostrophy) and is performed in terms of departures from the ICAO standard atmosphere for increased accuracy (Simmons and Chen 1991; see appendix A). The geopotential vertical interpolation and the temperature vertical interpolation are not exactly consistent with hydrostatic equilibrium, the emphasis having been put on accuracy. Nevertheless, consistent and accurate vertical interpolation could be devised, which may be important for intensive use of temperature information.

The vertical interpolation operators for SYNOP 10 m wind and 2 m temperature (T2m) match the model's surface-layer parametrization. The vertical gradients of the model variables vary strongly in the lowest part of the boundary layer, where flow changes are induced on very short time- and space-scales as a result of physical factors such as turbulence and terrain characteristics. The vertical interpolation operator for these data takes this into account, following Monin–Obukhov similarity theory. Results using such operators, which are inspired by the work of Geleyn (1988), have been presented by Cardinali *et al.* (1994). It was found that T2m data could not be used satisfactorily in the absence of surface skin temperature as part of the control variable, as unrealistic analysis increments appeared in the near-surface temperature gradients. The Monin–Obukhov based observation operator for 10 m wind on the other hand is used for SYNOP and DRIBU winds. Scatterometer 10 m winds, however, are used through a simple logarithmic relationship between lowest model-level wind (at approximately 32 m) and wind at 10 m.

$$u_{10\mathrm{m}} = u_{32\mathrm{m}} \frac{\ln(10/z_0)}{\ln(32/z_0)},$$

where z_0 denotes the surface roughness-length. Relative humidity is assumed constant in the lowest model-layer to evaluate its 2 m value. H has been expressed as the product of individual operators. The variational analysis procedure requires the gradient of the objective function with respect to the control variable. This computation makes use of the adjoint of the individual tangent–linear operators, applied in the reverse order. Operators also exist for SATEM precipitable water content and SSM/I total precipitable water data, but these data are currently not used operationally, nor in the experiments reported in Part III.

The details regarding observation operators for conventional data have been reported by Vasiljević *et al.* (1993). The TOVS observation operator has been described by Andersson *et al.* (1994) and the use of scatterometer data by Stoffelen and Anderson (1997).

(*b*) *Data usage*

Observation operators for all observation types that were used by OI have also been implemented in 3D-Var. In addition, 3D-Var uses TOVS cloud-cleared radiances and scatterometer ambiguous winds. Table 1 lists the observing systems currently used by 3D-Var in ECMWF's operational data assimilation. The table also indicates important restrictions on data usage and thinning of data. The TOVS data are further discussed in Part III. 3D-Var (like OI) uses the data from a six-hour time window centred at the analysis time. If there are multiple reports from the same fixed observing station within that time window, the data nearest the analysis time are selected for use in the analysis. Some thinning is applied for moving platforms reporting frequently.

(*c*) *Observation error covariances*

Initially, the OI standard deviations of observation errors were used in 3D-Var. However, the J_o diagnostics in 3D-Var (before minimization) showed that there was a misspecification of the observation-error standard-deviation σ_o for several data types. Particularly, SATOB and DRIBU winds had σ_o too large, and humidity errors were generally too small. Average J_o-ratios (defined as J_o/N_{obs} where N_{obs} is the number of observations) were computed from a 3D-Var assimilation, and compared with the statistical expectation of the J_o-ratio, which is $\langle J_o/N_{\mathrm{obs}}\rangle = (\sigma_o^2 + \sigma_b^2)/\sigma_o^2$. We can see that even with a perfect background, the ratio should not be less than one, and with $\sigma_b \approx \sigma_o$ it should be close to two. This led to modifications of the specified σ_o for a number of observation types. The observation errors currently used are tabulated in appendix B. The observation errors for TOVS radiance data were taken from Eyre *et al.* (1993), and are also listed in appendix B.

The observation error is assumed to be uncorrelated (i.e. the matrix $\mathbf{R}$ is diagonal) for all data except radiosonde geopotential data, TOVS radiance data and SATEM thicknesses. The radiosonde geopotential data are vertically correlated using a continuous covariance function $a \exp\{-b(x_1 - x_2)^2\}$, where a is 0.8, b a tuning constant close to one and x_1 and x_2 are transformation values, based on a sixth-degree polynomial in $\ln(p)$ of the two pressures involved. This is as previously used in ECMWF OI. The vertical correlation of SATEM thickness-data is as described by Kelly and Pailleux (1988). The horizontal correlation of SATEM and TOVS observation-errors are assumed to be Gaussian, and to have a length scale of 350 km. Only half the observation error is assumed to be correlated. The implementation has been described by Andersson *et al.* (1994), with arbitrary groupings of 120 pieces of data. The inter-channel correlation of radiance observation-error is assumed

TABLE 1. OBSERVATIONS USED BY 3D-VAR IN ECMWF'S OPERATIONAL ANALYSES IN 1996

Obs. type	Observed variable	Remark
SYNOP	Surface pressure	Not used if more than 800 m below model orography
	10 m wind	From SHIP and some tropical island stations
	2 m relative humidity	Not used if more than 50 m below model orography
AIREP	Wind	Thinned to approximately 200 km along flight
	Temperature	Thinned to approximately 200 km along flight
SATOB	Wind	METEOSAT, GOES and GMS (not over land in northern hemisphere nor over land for low-level winds)
DRIBU	Surface pressure	
	10 m wind	
TEMP	Geopotential height	Maximum 15 levels (priority to standard levels)
	Wind	Maximum 15 levels (priority to standard levels)
	Specific humidity	Below 300 hPa
PILOT	Wind	Maximum 15 levels (priority to standard levels)
TOVS	Cloud-cleared radiances	Tropospheric channels over oceans only south of 70°N
	1D-Var retrieval	North of 70°N
SATEM	NESDIS retrieval	Above 100 hPa poleward of 20°.
PAOB	Surface pressure	South of 20°S
SCAT	Ambiguous winds	If both observed and model wind speeds are less than 20 m s^{-1}

to be zero. The horizontal correlation of SATEM and TOVS observation-errors ceased to be accounted for in ECMWF operations at the end of September 1996. The effect on analysis and forecast accuracy is small.

5. PRACTICAL IMPLEMENTATION

(*a*) *Incremental formulation*

As mentioned earlier in the description of 3D-Var, the formulation used is incremental (Courtier *et al.* (1994)). The cost is then comparable to the cost of OI. This requires that two different resolutions are used for the comparison with observations (T213L31) and minimization (T63L31). Three different steps are performed: the comparison of the background with the observations at high resolution to compute the innovation vectors, the minimization at low resolution to produce low-resolution analysis increments, and finally the comparison of the analysis field with observations.

A truncation operator allows one to go from high-resolution fields to low resolution. For the spectral upper-air fields, this is simply a truncation at wave-number 63. The grid-point fields are first transformed to spectral space, then truncated, and transformed back to grid-point space. In that way, spectral upper-air fields and surface grid-point fields used in the minimization are consistent.

The analysis field is the sum of the background and of the pseudo-inverse of the truncation operator applied to the low-resolution increments. This pseudo-inverse consists in filling in the spectral wave-numbers greater than 63 by 0, and then applying adiabatic nonlinear normal mode initialization (NNMI) to the increments in order to adjust the high wave-numbers to the orography at high resolution. This is performed by evaluating

$$\mathbf{x}_a(\mathrm{HR}) = \mathbf{x}_b(\mathrm{HR}) + \mathrm{NNMI}(\mathbf{x}_b(\mathrm{HR}) + \delta\mathbf{x}_a(\mathrm{LR})) - \mathrm{NNMI}(\mathbf{x}_b(\mathrm{HR}))$$

where HR and LR stands for high and low resolution respectively. This is necessary since the analysis obtained at the end of the minimization procedure is balanced but for an orography truncated at wave-number 63.

As far as the humidity is concerned, the control variable used in the minimization is specific humidity in spectral space. There is no constraint forcing the minimization to produce positive and non-supersaturated values for this quantity. However, before the computation of observation departures in the minimization stage, grid-point values are replaced by $f(q, q_{\mathrm{sat}})$ where f is a differentiable function such that it results in positive and not supersaturated humidity values. The high-resolution analysis is modified by resetting negative humidities to zero and supersaturated values to saturated values.

(*b*) *Preconditioning*

As already explained in subsection 3(a)(ii), one generally uses the Hessian of the background term J_b to precondition the problem. One implementation is to use the change of variable $\mathbf{z} = \mathbf{B}^{-1/2}\delta\mathbf{x}$. In practice, this was not found to be practical in the current formulation of 3D-Var where the J_b computation involves splitting the increment $\delta\mathbf{x}$ into three parts: the Rossby part, the gravity-wave part and a univariate part.

(*c*) *Minimization*

The minimization problem involved in this 3D-Var can be considered as large-scale, since the number of degrees of freedom in the control variable is slightly over 500 000. An efficient descent algorithm was provided by the Institut de Recherche en Informatique et Automatique (INRIA), France. It is a variable-storage quasi-Newton algorithm described by Gilbert and Lemaréchal (1989). The method uses the available in-core memory to update an approximation of the Hessian of the cost function. In practice, ten updates of this Hessian matrix are used. The approximation is modified during the minimization by deleting information from the oldest gradient and inserting information from the most recent one. The number of iterations is limited to 70. Only very small adjustments of the analysis occur during the second half of the minimization. On the whole, the cost function is typically divided by a factor of two and the norm of the gradient by a factor of twenty. The approximation of the Hessian computed during a given analysis 24 hours earlier is used as a first estimate for the corresponding current analysis.

6. Numerical illustrations

Single-observation analysis is used to exhibit the structure functions as specified in variational assimilation since the analysis increments are then proportional to the covariances of the background errors in model space, with the background errors in the observation space for the observation considered. This has been explained in detail by Thépaut *et al.* (1996) who applied this approach to compare the statistical models of 3D-Var and 4D-Var for the height field in an example of mid-latitude cyclogenesis. We follow the same approach to illustrate the multivariate aspects of 3D-Var.

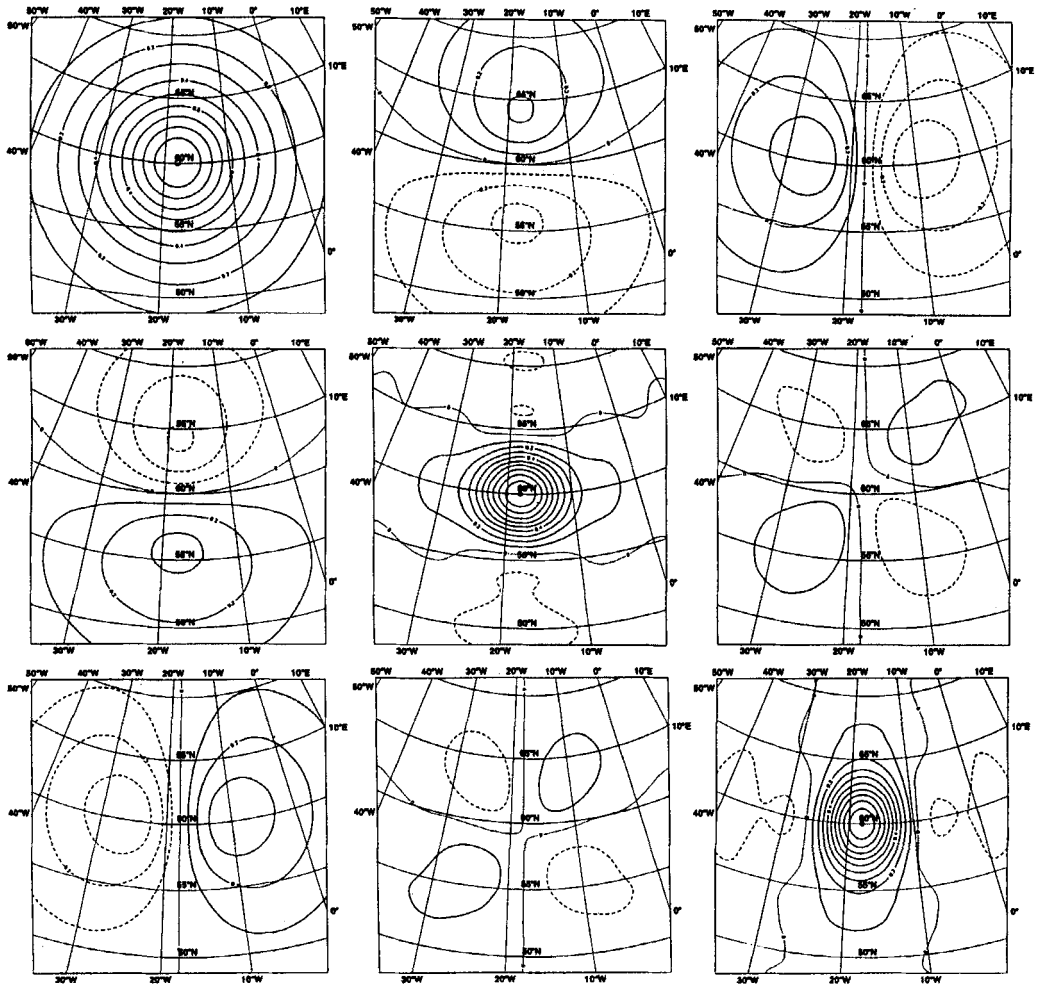

Figure 1. 500 hPa 3D-Var structure functions at 60°N. The first row depicts the z:z (left), z:u (middle) and z:v (right) correlations, the second row depicts the u:z (left), u:u (middle) and u:v (right) correlations, and the third row depicts the v:z (left), v:u (middle) and v:v (right) correlations.

Figures 1 and 2 show the results of such single-observation experiments with the background error statistics as obtained in Part II. The observations are located at 60°N and at the equator, respectively. Horizontally-constant standard deviations of background errors have been specified and J_c was not used. The three rows are for observation of 500 hPa geopotential height (top), u-component (middle) and v-component (bottom) of the 500 hPa wind. The three columns show the response in terms of height, u and v, respectively. The results have been expressed in terms of correlations, computed from the analysis increments by normalization with the effective standard deviations of background errors (following Thépaut *et al.* (1996) in their 3D-Var results). As expected, the results in mid-latitudes (Fig. 1) are near-geostrophic, and they are in close agreement with those shown in similar diagrams by Mitchell *et al.* (1990), Daley (1991) and Bartello and Mitchell (1992), for example. The unusual shape of the negative lobes of the u:u correlations was also found by Bartello and Mitchell. Our results indicate that the u:z and z:u correlations have slightly higher magnitude to the south than to the north of the observation. This is due to

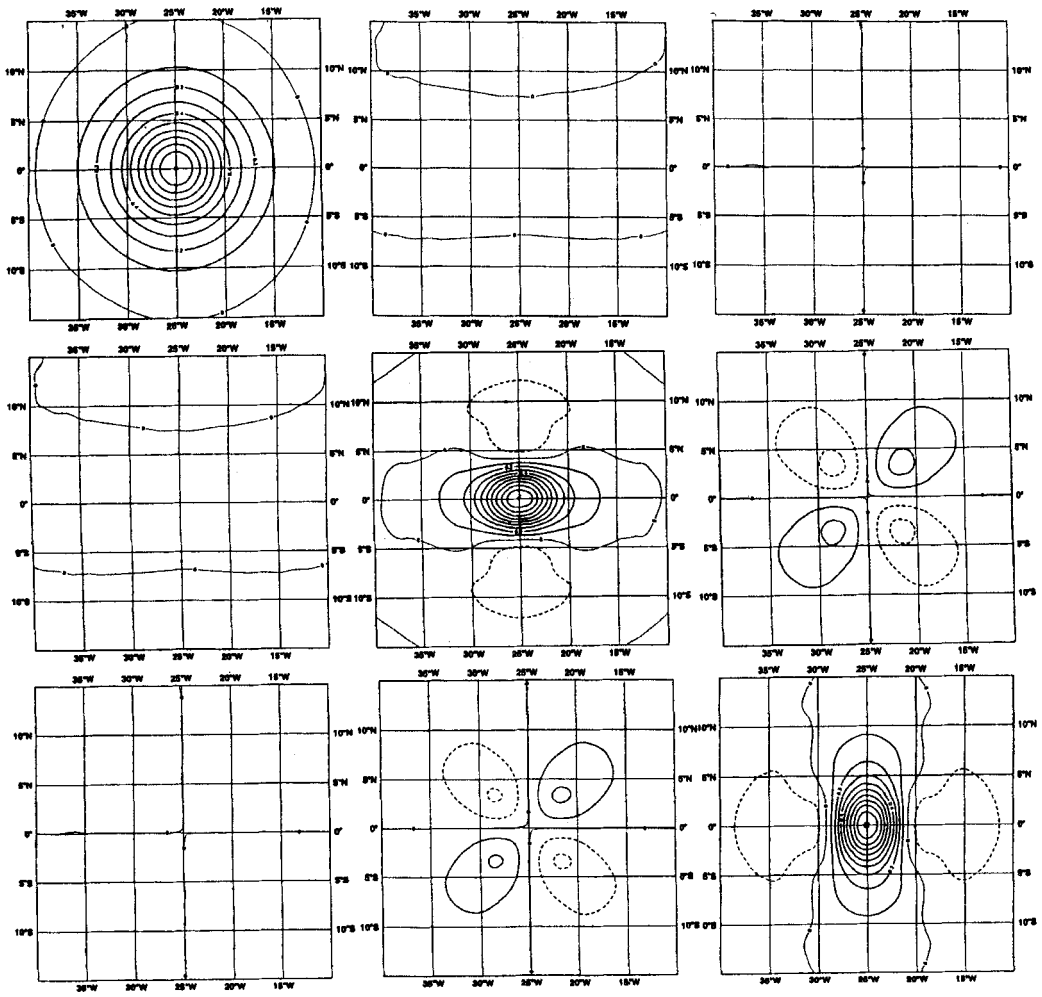

Figure 2. Same as Fig. 1 but at the equator.

the variation in the Coriolis parameter f. The results on the equator (Fig. 2) show that the mass–wind coupling is very weak as expected from the univariate tropical formulation, with the wind increments mostly non-divergent.

Figure 3 shows two cross-sections of covariances, in a vertical plane along 60°N: (a) v:v; (b) $v : T$. It can be seen that, in a vertical column, the v:v covariances diminish more rapidly downward than upward which confirms the result of Bartello and Mitchell (1992). This is an expression of the non-separability between the horizontal and vertical directions of the background-error structure-functions, as will be dicussed further in Part II. As a consequence, v:T covariances (Fig. 3(b)) tend to slope upwards, away from the observation.

7. Conclusion

The general conclusion of this three-part paper is deferred to Part III. In the present part we discuss only the specifics of the formulation. We have shown that it is possible within

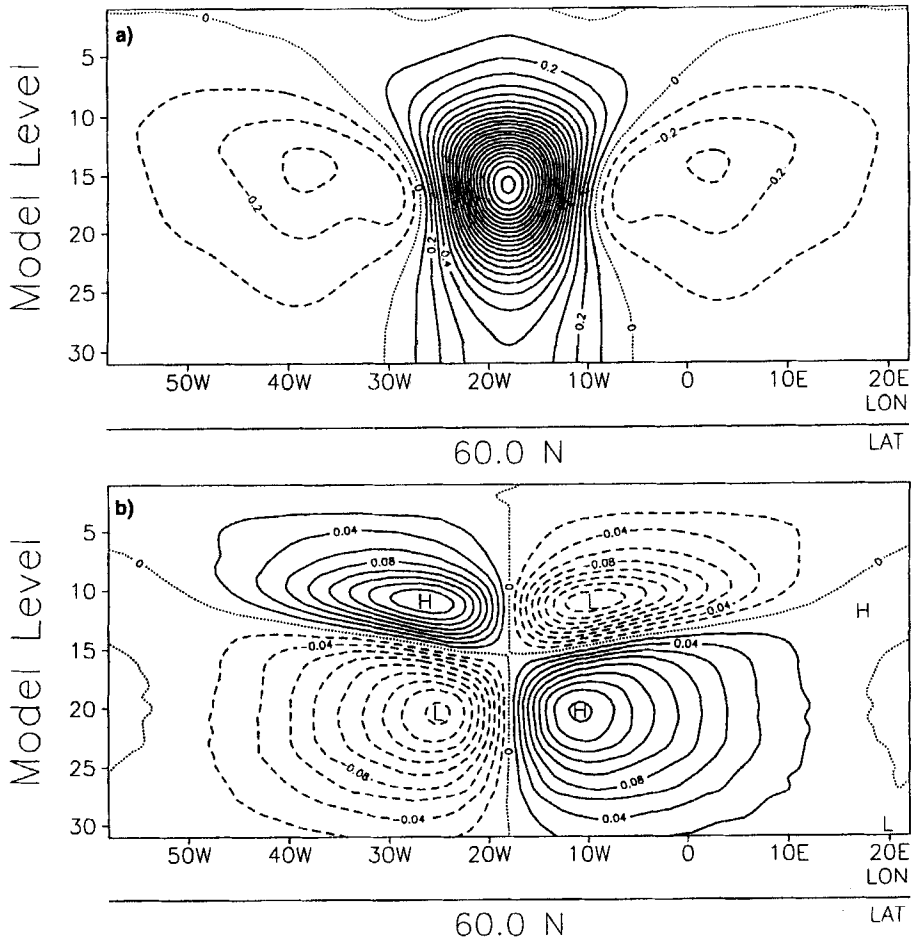

Figure 3. Covariances in a north–south vertical plane: (a) v:v and (b) v:T.

3D-Var to account for most features of optimal interpolation. With respect to the ECMWF OI, the 3D-Var statistical model has been refined to incorporate non-separability in a way similar to that suggested by Phillips (1986). The multivariate formulation makes use of the Hough functions which paves the way toward a formulation which would produce analysis increments tangent to the slow manifold. Nevertheless, the present multivariate formulation does not have an explicit square root. This is a weakness for the preconditioning, which assumes, for example, that the vertical correlations of divergence and vorticity are kept the same for a given total wave-number. In addition, the simplified Kalman filter as proposed by Courtier (1994) requires a square-root formulation. In the present formulation, J_b is evaluated in spectral space; it would be possible to evaluate it in Hough-modes space, since recent ECMWF developments towards a simplified Kalman-filter imply smoother background-error standard-deviations than those of OI.

No progress has been made in the formulation of 3D-Var in the tropics as compared to OI. The horizontally variable vertical structure-functions, when implemented, may allow reconsideration of the balance aspects. 3D-Var analyses the fields of mass, wind and humidity simultaneously. However, in the present formulation, the humidity field remains

largely decoupled from the others. A significant step forward may be expected from a proper coupling of the mass and wind fields with the humidity field.

Significant difficulties have been encountered with the treatment of tidal waves. With the six-hour cycles used in 3D-Var, any mis-specification of the semi-diurnal tidal wave can be interpreted as a $2\Delta t = 12$ hours oscillation of the analysis increments. The difficulty is three-fold: such an oscillation is a non-dissipated solution of the model equations; any biases in the model or the conventional network will alias the tides; any biases in the two TOVS satellites, one with respect to the other, will alias the semi-diurnal tide as they are in quadrature. The initialization of the large-scale component of the analysis increments cured the problem to a large extent. Nevertheless, this remains a challenge for the future improvement of the data-assimilation system.

Two potential weaknesses of 3D-Var related to the vertical coordinate have been exhibited. The performance of 3D-Var will have to be carfully monitored in the vicinity of orography, and specific remedies may have to be devised.

The incremental formulation may be extended to the vertical, following Gauthier *et al.* (1996) who devised a 3D-Var formulation in pressure coordinates. In particular, this may be a significant improvement if the isotropy hypothesis is better satisfied in the free atmosphere along, for example, potential-temperature surfaces. For the horizontal aspects, a computational sphere may be introduced with the isotropy being applied over this sphere. The mapping between the geographical sphere and the computational sphere could rely on the semi-geostrophic coordinates over the plane, as proposed by Desroziers and Lafore (1993).

Acknowledgements

The first author would like to thank P. Gauthier for very fruitful discussions in Montreal in July 1996. The development of 3D-Var greatly benefited from the ARPEGE–IFS collaboration, particularly from the contribution of Jean-Noël Thépaut, Patrick Moll, Philippe Caille, Per Undén and the ECMWF data-division.

Fréderic Delsol and Dave Burridge believed in the variational project in 1986–1987; Michel Jarraud, Jean-François Geleyn and Michel Rochas, on the continent, realized that this dream (or probably nightmare for some!) might be transformed into a palpable project.

The 'Groupement de Recherche sur les Méthodes variationnelles en Météorologie et Océanographie' funded by the 'Centre National de la Recherche Scientifique' and Météo-France has supported the variational project financially from 1988 to 1996, with the Director, Olivier Talagrand, being a staunch supporter, particularly as the Chairman of the ECMWF Scientific Advisory Committee.

Adrian Simmons and Peter Houtekamer are acknowledged for their valuable comments on a first version of the manuscript. John Derber highlighted several weaknesses of this formulation.

Appendix A

The observation operator for geopotential height

The geopotential at a given pressure p is computed by integrating the hydrostatic equation analytically using the International Civil Aviation Organization (ICAO) temperature profile and vertically interpolating $\Delta\phi$, the difference between the model-level geopotential and the ICAO geopotential. The ICAO temperature profile is defined as

$$T_{\mathrm{ICAO}} = T_0 - \frac{\Lambda}{g}\phi_{\mathrm{ICAO}}\,, \tag{A.1}$$

where T_0 is 288 K, ϕ_{ICAO} is the geopotential above 1013.25 hPa and Λ is 0.0065 K m^{-1} in the ICAO troposphere and 0 K m^{-1} in the ICAO stratosphere. The ICAO tropopause is defined as the level where the ICAO temperature has reached 216.5 K. Using this temperature profile and integrating the hydrostatic equation provides T_{ICAO} and the geopotential ϕ_{ICAO} as a function of pressure. We may then evaluate the geopotential $\phi(p)$ at any pressure p following

$$\phi(p) - \phi_s = \phi_{\mathrm{ICAO}}(p) - \phi_{\mathrm{ICAO}}(p_s) + \Delta\phi\,, \tag{A.2}$$

where p_s is the model surface-pressure and ϕ_s the model orography. $\Delta\phi$ is obtained by vertical interpolation from the full model-level values $\Delta\phi_k$. The interpolation is linear in $\ln(p)$ up to the second model level and quadratic in $\ln(p)$ for levels above it. The full model-level values are obtained integrating the discretized hydrostatic equation following Simmons and Burridge (1981)

$$\Delta\phi_k = \sum_{j=N_{\mathrm{lev}}}^{k+1} R_d(T_{v_j} - T_{\mathrm{ICAO}_j})\ln\left(\frac{p_{j+1/2}}{p_{j-1/2}}\right) + \alpha_k R_d(T_{v_k} - T_{\mathrm{ICAO}_k}) \tag{A.3}$$

with

$$\alpha_k = 1 - \frac{p_{k-1/2}}{p_{k+1/2} - p_{k-1/2}}\ln\left(\frac{p_{k+1/2}}{p_{k-1/2}}\right)$$

for $k > 1$ and $\alpha_1 = \ln(2)$.

Appendix B

Observation errors

The observation-error standard deviations currently assumed in the ECMWF 3D-Var are given in the tables, for upper-air data and SYNOP heights (Table B.1), surface data (Table B.2), TOVS radiance data (Table B.3) and SATEM retrieved thicknesses (Table B.4). They include the instrumental error and the representativeness error, and thus for

TABLE B.1. Observation errors, upper air

Pressure (hPa)	TEMP u/v (m s^{-1})	SATOB u/v (m s^{-1})	AIREP u/v (m s^{-1})	PILOT u/v (m s^{-1})	AIREP T (K)	TEMP Z (m)	SYNOP Z (m)
1000	2.3	2.0	2.5	2.3	1.4	4.3	7.0
850	2.3	2.0	2.5	2.3	1.4	4.4	8.0
700	2.5	2.0	3.0	2.5	1.2	5.2	8.6
500	3.0	3.5	3.5	3.0	1.2	8.4	12.1
400	3.5	4.3	4.0	3.5	1.2	9.8	14.9
300	3.7	5.0	4.0	3.7	1.3	10.7	
250	3.5	5.0	4.0	3.5	1.3	11.8	
200	3.5	5.0	4.0	3.5	1.4	13.2	
150	3.4	5.0	4.0	3.4	1.4	15.2	
100	3.3	5.0	4.0	3.3	1.4	18.1	
70	3.2	5.0	4.0	3.2	1.5	19.5	
50	3.2	5.0	4.0	3.2	1.6	22.5	
30	3.3	5.0	4.0	3.3	1.8	25.0	
20	3.6	5.0	4.0	3.6	2.0	32.0	
10	4.5	5.7	4.0	4.5	2.2	40.0	

TABLE B.2. OBSERVATION ERRORS, SURFACE

PAOB Z (m)	DRIBU Z (m)	SHIP Z (m)	SYNOP u/v ($m\ s^{-1}$)	DRIBU u/v ($m\ s^{-1}$)	SYNOP T (K)	SHIP T (K)	DRIBU T (K)
24.0	11.5	14.0	3.0	2.4	2.0	1.8	1.8

TABLE B.3. OBSERVATION ERRORS, TOVS RADIANCES (K)

Channel	Error	Channel	Error	Channel	Error
HIRS		HIRS		MSU	
1	0.7	10	0.8	2	0.3
2	0.35	11	1.1	3	0.22
3	0.3	12	1.5	4	0.25
4	0.2	13	0.5		
5	0.3	14	0.35		
6	0.4	15	0.3		
7	0.6				

TABLE B.4. OBSERVATION ERRORS, SATEM THICKNESSES (m)

Layer (hPa)	NESDIS clear	NESDIS p. cloudy cloudy	1D-Var clear	1D-Var p. cloudy	1D-Var cloudy
1000–700	27.0	33.0	8.81	9.53	12.77
700–500	17.0	20.0	7.70	8.39	11.70
500–300	25.0	31.0	13.23	13.61	18.31
300–100	53.0	56.0	19.62	20.22	22.86
100–70	33.0	33.0			
70–50	36.0	36.0			
50–10	74.0	74.0			

the TOVS data they include the forward model error. The TOVS observation errors in partly cloudy spots are inflated to 0.5, 0.9, 1.3, 0.9, 0.6 and 0.40 for channels HIRS-6, 7, 10, 13, 14 and 15, to account for any residual cloud contamination. Furthermore, the radiance observation errors are multiplied by a factor 1.5 when used in 3D-Var, because the data are sensitive to quantities such as temperature above the model top and surface temperature, which are currently not part of the 3D-Var control variable. The relative humidity errors are specified according to a regression relation $\sigma_o = -0.0015T + 0.54$, for 240 K $< T <$ 320 K, obtained from a statistical analysis of their dependence with respect to temperature. Outside this interval, σ_o takes the values 0.18 and 0.06 respectively. The standard deviation of error for the specific humidity is then deduced according to the tangent–linear relationship which relates specific and relative humidity, the linearization being performed in the vicinity of the observed values of temperature and humidity.

REFERENCES

Abramowitz, M. and Stegun, I. 1965 *Handbook of mathematical functions.* Dover publications Inc., New York

Andersson, E., Hollingsworth, A., Kelly, G., Lönnberg, P., Pailleux, J. and Zhang, Z. 1991 Global observing system experiments on operational statistical retrievals of satellite sounding data. *Mon. Weather Rev.*, **119**, 1851–1864

Andersson, E., Pailleux, J., Thépaut, J.-N., Eyre, J. R., McNally, A. P., Kelly, G. A. and Courtier, P. 1994 Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120**, 627–653

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljević, D., Branković, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N. and Viterbo, P. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, 1831–1860

Bartello, P. and Mitchell, H. L. 1992 A continuous three-dimensional model of short-range forecast error covariance. *Tellus*, **44A**, 217–235

Boer, G. J. 1983 Homogeneous and isotropic turbulence on the sphere. *J. Atmos. Sci.*, **40**, 154–163

Bourke, W. 1972 An efficient one level primitive equation spectral model. *Mon. Weather Rev.*, **100**, 683–689

Bouttier, F. 1994 A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Weather Rev.*, **122**, 2376–2390

Bouttier, F., Derber, J. and Fisher, M. 1997 'The 1997 revision of the J_b term in 3D/4D-Var'. ECMWF Tech. Memo. 238 (available from ECMWF, Reading, UK)

Cardinali, C., Andersson, E., Viterbo, P., Thépaut, J.-N. and Vasiljević, D. 1994 'Use of conventional surface observations in three-dimensional variational data assimilation'. ECMWF Tech. Memo. 205 (available from ECMWF, Reading, UK)

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. 1998 Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Weather Rev.*, **126**. In press

Courtier, P. 1987 'Application du contrôle optimal à la prévision numérique en météorologie'. Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris, France

1994 'Introduction to numerical weather prediction data assimilation methods'. Pp. 189–208 in proceedings of ECMWF seminar on developments in the use of satellite data in numerical weather prediction. 6–10 September 1993 (available from ECMWF, Reading, UK)

1997 Dual formulation of four dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, **123**, 2449–2461

Courtier, P. and Naughton, M. 1994 A pole problem in the reduced Gaussian grid. *Q. J. R. Meteorol. Soc.*, **120**, 1389–1407

Courtier, P. and Talagrand, O. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Q. J. R. Meteorol. Soc.*, **113**, 1329–1347

1990 Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531–549

Courtier, P., Andersson, E., Heckley, W., Kelly, G., Pailleux, J., Rabier, F., Thépaut, J.-N., Undén, P., Vasiljević, D., Cardinali, C., Eyre, J., Hamrud, M., Haseler, J., Hollingsworth, A., McNally, A. and Stoffelen, A. 1993 'Variational assimilation at ECMWF'. ECMWF Tech. Memo. 194 (available from ECMWF, Reading, UK)

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

Daley, R. 1991 *Atmospheric data analysis*. Cambridge atmospheric and space sciences series, Cambridge University Press

Dee, D. P. 1995 On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.*, **123**, 1128–1145

Derber, J., Parrish, D. F. and Loord, S. J. 1991 The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, **6**, 538–547

Desroziers, G. and Lafore, J.-P. 1993 Coordinate transformation for objective frontal analysis. *Mon. Weather Rev.*, **121**, 1531–1553

Eyre, J. R. 1989 Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. *Q. J. R. Meteorol. Soc.*, **115**, 1001–1037

Eyre, J. R. and Lorenc, A. C. 1989 Direct use of satellite sounding radiances in numerical weather prediction. *Meteorol. Mag.*, **118**, 13–16

Eyre, J. R., Kelly, G. A., McNally, A. P., Andersson, E. and Persson, A. 1993 Assimilation of TOVS radiance information through one-dimensional variational analysis. *Q. J. R. Meteorol. Soc.*, **119**, 1427–1463

Flobert, J-F., Andersson, E., Chédin, A., Hollingsworth, A., Kelly, G., Pailleux, J. and Scott, N. A. 1991 Global data assimilation and forecast experiments using the Improved Initialization Inversion method for satellite soundings. *Mon. Weather Rev.*, **119**, 1881–1914

Gandin, L. S. 1963 *Objective Analysis of Meteorological Fields.* Leningrad, Gidromet.; Jerusalem, Israel Program for Scientific Translations, 1965

Gaspari, G. and Cohn, S. E. 1999 Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, in press

Gauthier, P., Fillion, L., Koclas, P. and Charette, C. 1996 'Implementation of a 3D variational analysis at the Canadian Meteorological Centre'. In proceedings of XIth AMS Conference on Numerical Weather Prediction, Norfolk, Virginia, 19–23 August 1996, AMS, Boston, USA

Geleyn, J. F. 1988 Interpolation of wind, temperature and humidity values from the model levels to the height of meaurement. *Tellus*, **40**, 347–351

Gel'Fand, I. M. and Vilenkin, N. Ya. 1964 *Generalized Functions, Vol. 4: Applications of Harmonic Analysis.* Academic Press, New York and London

Gilbert, J. C. and Lemaréchal, C. 1989 Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Prog.*, **B25**, 407–435

Heckley, W., Courtier, P., Pailleux, J. and Andersson, E. 1993 'The ECMWF variational analysis: general formulation and use of background information'. Pp. 49–94 in proceedings of ECMWF workshop on variational assimilation with special emphasis on three-dimensional aspects, 9–12 November 1992 (available from ECMWF, Reading, UK)

Hollingsworth, A. 1987 'Objective analysis for numerical weather prediction'. Pp. 11–60 in special volume *J. Meteorol. Soc. Japan, Short and Medium Range Numerical Weather Prediction*, Matsuno T. Ed.

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

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. 1997 Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Japan*, **75**, 181–189

Kelly, G. and Pailleux, J. 1988 'Use of satellite vertical sounder data in the ECMWF analysis system'. ECMWF Tech. Memo. 143 (available from the European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK)

Kelly, G., Andersson, E., Hollingsworth, A., Lönnberg, P., Pailleux, J. and Zhang, Z. 1991 Quality control of operational physical retrievals of satellite sounding data. *Mon. Weather Rev.*, **119**, 1866–1880

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

Lorenc, A. C. 1981 A global three-dimensional multivariate statistical interpolation scheme. *Mon. Weather Rev.*, **109**, 701–721

1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194

1988 Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, **114**, 205–240

1992 Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.*, **118**, 569–591

Lorenc, A. C. 1995 'Development of an operational variational assimilation scheme'. Pp. 415–420 in Proceedings of WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Tokyo, Japan

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

Müller, R. 1989 A note on the relation between the 'traditional approximation' and the metric of the primitive equations. *Tellus*, **41A,** 175–178

Pailleux, J. 1990 'A global variational assimilation scheme and its application for using TOVS radiances'. Pp. 325–328 in Proceedings of WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography. Clermont-Ferrand, France

Parrish, D. F. and Derber, J. C. 1992 The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.*, **120,** 1747–1763

Parrish, D. F., Derber, J. C., Purser, J., Wu, W. and Pu, Z. 1995 'The NMC global analysis system: Recent developments and future plans'. Pp. 403–407 in Proceedings of WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Tokyo, Japan

Phillips, N. A. 1973 'Principles of large scale numerical weather prediction'. Pp. 1–96 in *Dynamic meteorology.*, Ed. P. Morel. Reidel, Dordrecht, Netherlands

1986 The spatial statistics of random geostrophic modes and first-guess errors. *Tellus*, **38A,** 314–332

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and Bouttier, F. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124,** 1809–1829

Rutherford, I. 1972 Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29,** 809–815

Shaw, D.B., Lönnberg, P., Hollingsworth, A. and Undén, P. 1987 Data assimilation: The 1984/1985 revisions of the ECMWF assimilation system. *Q. J. R. Meteorol. Soc.*, **113,** 533–566

Simmons, A. J. and Burridge, D. 1981 An energy and angular momentum conserving vertical finite difference scheme and hybrid coordinate. *Mon. Weather Rev.*, **109,** 758–766

Simmons, A. J. and Chen, J. 1991 The calculation of geopotential and the pressure gradient in the ECMWF atmospheric model: Influence on the simulation of the polar atmosphere and on temperature analyses. *Q. J. R. Meteorol. Soc.*, **117,** 29–58

Stoffelen, A. and Anderson, D. 1997 Ambiguity removal and assimilation of scatterometer data. *Q. J. R. Meteorol. Soc.*, **123,** 491–518

Talagrand, O. and Courtier, P. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc.*, **113,** 1311–1328

Thépaut, J.-N. and Moll, P. 1990 Variational inversion of simulated TOVS radiances using the adjoint technique. *Q. J. R. Meteorol. Soc.*, **116,** 1425–1448

Thépaut, J.-N., Courtier, P., Belaud, G. and Lemaître, G. 1996 Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Q. J. R. Meteorol. Soc.*, **122,** 535–561

Undén, P. 1989 Tropical data assimilation and analysis of divergence. *Mon. Weather Rev.*, **117,** 2495–2517

Vasiljević, D., Cardinali, C. and Undén, P. 1993 'ECMWF 3D-Variational assimilation of conventional observations'. In proceedings of the ECMWF workshop on 'Variational assimilation with emphasis on three-dimensional aspects', 9–12 November 1992 (available from ECMWF, Reading, UK)

Wergen, W. 1988 Diabatic non-linear normal mode initialisation for a spectral model. *Beitr. Phys. Atmos.*, **61,** 274–302

Wunsch, C. and Stammer, D. 1995 The global frequency–wavenumber spectrum of oceanic variability estimated from TOPEX/POSEIDON altimetric measurements. *J. Geophys. Res.*, **100,** 24895–24910

Yaglom, A. M. 1987 *Correlation theory of stationary and related random functions, Vol. I: Basic results.* Springer-Verlag, New York, USA