TECHNICAL MEMORANDUM

194

Variational assimilation at ECMWF

P. Courtier, E. Andersson, W.A. Heckley,
G. Kelly, J. Pailleux, F. Rabier,
J N Thépaut, P. Undén,
D.Vasiljević, C. Cardinali,
J. Eyre, M. Hamrud, J. Haseler,
A. Hollingsworth, A.P. McNally,
and A. Stoffelen

Research Department

September 1993

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Variational Assimilation at ECMWF

Document presented at the 21st Session of the Scientific Advisory Committee

1. INTRODUCTION

This paper summarises the progress made at ECMWF in variational data assimilation since SAC(91)4 (also available as *Pailleux et al*, 1991). Two years ago we presented a feasibility study of the major components of 3D-Var and early results of 4D-Var. Since then the scientific validation in analysis and assimilation mode has started and some reorientation of the project took place in the light of these results.

In the following we shall describe the current 3D-Var and 4D-Var formulation. Section 2 presents the theoretical aspects and introduces the incremental approach. Section 3 describes the background error term formulation and the specification of the statistics as well as the control of the gravity waves. Section 4 presents the observation term. The following 2 sections describe the major results obtained with 3D-Var and 4D-Var. Finally section 7 describes further lines of research.

2. THEORETICAL ASPECTS

Let us denote by $M(t_2,t_1)$ the model integrated from time t_1 to t_2; it is used to carry in time the state of the atmosphere x:

$$x(t_2) = M(t_2,t_1)\, x(t_1) \tag{1}$$

Suppose we intend to perform a 4D-Var assimilation over a period (t_o,t_o+T-t_N), (to fix ideas we let $T = 24 hours$). N is the number of time steps necessary to integrate the model from time t_o to time $t_o + T$.

Over this time interval, observations y_i are available at each time t_i. We assume that all observations available between times $t_{(i-½)}$ and $t_{(i+½)}$ are valid for time t_i. This is not a serious limitation since we are already below the time scales resolved by the model. The observation y_i is linked to the model state variable $x(t_i)$ by the observation operator H_i

$$y_i = H_i\, x(t_i) + \varepsilon_i \tag{2}$$

Eq. (2) defines the observation error ε_i of covariance matrix O_i, which consists of the sum of the measurement errors and the representativeness errors (*Lorenc*, 1986). As detailed by *Lorenc* (1986)

4D-Var consists of the minimization problem

$$\mathcal{P}_{4D}\text{: minimise } J(x(t_o)) = \tfrac{1}{2}(x(t_o)-x_b)^t\, B^{-1}(x(t_o)-x_b) + \tfrac{1}{2}\sum_{i=0}^{N} (H_i x(t_i)-y_i)^t\, O_i^{-1}(H_i x(t_i)-y_i) \tag{3}$$

with $x(t_i) = M(t_i,t_o)x(t_o)$

x_b is the background information valid for time t_o which summarises all the information used before time t_o, and B is the error covariance matrix of x_b.

A classical result, assuming a perfect model and linearity of H_i and M, is that if $x^*(t_o)$ is the result of $\mathcal{P}_{4D}$, then $x^*(t_N) = M(t_N,t_o)\; x^*(t_o)$ can also be obtained applying the Kalman filter to the same statistical estimation problem (*Jazwinski*, 1970; *Lorenc*, 1986; see *Thépaut and Courtier*, 1991 or *Rabier et al.*, 1993 for a detailed presentation using the same notations as here). In meteorological applications, however, H_i and M are weakly nonlinear: assuming that the tangent linear operators $\mathfrak{M}$ and H_i' of respectively M and H_i satisfy, to acceptable accuracy the relations

$$\begin{aligned} M(t_i,t_o)\;(x(t_o)+\delta x(t_o)) &\approx M(t_i,t_o)\; x(t_o) + \mathfrak{M}(t_i,t_o)\delta x(t_o) \\ H_i(x(t_i) + \delta x(t_i)) &\approx H_i x(t_i) + H_i'\cdot\delta x(t_i) \end{aligned} \tag{4}$$

for meteorological order of magnitude of the perturbation $\delta x(t_o)$, then the 4D-Var problem $\mathcal{P}_{4D}$ is equivalent to the so-called extended Kalman filter (see previous references). This consists of two steps (^{f} and ^{a} denote respectively forecast and analysis)

THE FORECAST STEP

$$x^f(t_{i+1}) = M(t_{i+1},t_i)x^a(t_i) \tag{5a}$$

$$B^f(t_{i+1}) = \mathfrak{M}(t_{i+1},t_i)\; B^a(t_i)\; \mathfrak{M}^t(t_{i+1},t_i) \tag{5b}$$

where the state vector is advanced by the full model (Eq. 5a) and the forecast error covariance is advanced by the tangent linear model (Eq. 5b).

THE ANALYSIS STEP

$$x^a(t_i) = x^f(t_i) + K_i(y_i-H_i x(t_i)) \tag{6a}$$

$$B^a(t_i) = (I-K_i H_i')\; B^f(t_i) \tag{6b}$$

where $\boldsymbol{K}_i = \boldsymbol{B}^f(t_i)\boldsymbol{H}'^t_i[\boldsymbol{H}'_i\boldsymbol{B}^f(t_i)\,\boldsymbol{H}'^t_i + \boldsymbol{O}_i]^{-1}$ (6c)

is given by the minimum variance optimality condition.

For the equivalence to hold, the initial conditions $\boldsymbol{x}^f(t_o)$ and $\boldsymbol{P}^f(t_o)$ of the Kalman filter are

$$\boldsymbol{x}^f(t_o) = \boldsymbol{x}_b$$
$$\boldsymbol{B}^f(t_o) = \boldsymbol{B}$$

In current operational practice Eq. (5a) is treated exactly and Eqs. (6a) and (6c) are solved to a good accuracy; however Eq. (5b) is very crudely approximated with the so-called structure functions. These are the specified spatial error correlations which are kept constant in time in current practice, while the analysis variances are amplified with a very simple rule. By contrast, 4D-Var implicitly uses flow-dependent structure functions in Eq. (5b) (*Thépaut et al.*, 1993a), so that 4D-Var is a scientific improvement on the current operational implementation. Moreover, 4D-Var is also an algorithmic improvement on the Kalman filter (5, 6) where the Eq. (5b) has to be solved explicitly.

There are three main weaknesses in the 4D-Var implementation. First the model is assumed to be perfect: in Eq. (5b) no source terms $\boldsymbol{Q}$ are present; some negative effects of this approximation can be found in (*Wergen*, 1992). Secondly, if the unknown filter is obviously sequential, cycling 4D-Var requires approximations on the specification of $\boldsymbol{B}$. Thirdly, we do not have access to the analysis error covariance $\boldsymbol{B}^a(t_N)$.

Here we suggest that if (5b) is approximate anyway, it is not scientifically worthwhile solving it exactly. In other words, it may be scientifically acceptable to replace $\boldsymbol{\Re}$ by an approximate tangent linear model in (5b) provided this approximation is smaller than the approximation of neglecting the model error source term $\boldsymbol{Q}$.

Let us assume from now on that $\boldsymbol{\Re}$ is any linear operator, for which we will later stipulate the link with the model $\boldsymbol{M}$. We define the 4D-Var problem:

$$\wp'_{4D}\text{: minimise } J(\delta x(t_o)) = \tfrac{1}{2}\delta x(t_o)^t\, B^{-1}\,\delta x(t_o) + \tfrac{1}{2}\sum_{i=0}^{N}\,(H_i x(t_i)-y_i)^t\, O_i^{-1}(H_i x(t_i)-y_i) \qquad (7)$$

with $x(t_i) = M(t_i,t_o)\,x_b + \Re(t_i,t_o)\,\delta x(t_o)$

REMARK 1

If $\mathfrak{R}$ is the tangent linear model, $\wp'_{4D}$ and $\wp_{4D}$ are equivalent to within the accuracy of the tangent linear approximation.

REMARK 2

If $\mathfrak{R}$ is any linear operator which we assume describes exactly the forecast error evolution, $\wp'_{4D}$ leads to a statistically optimal result, the same result as the Kalman filter described by Eqs. (5) and (6).

REMARK 3

$\wp'_{4D}$ is better than $\wp_{4D}$ as far as an operational implementation is concerned since we keep the original model M for propagating in time the state of the atmosphere, but use an approximate propagation in time of the errors, thus introducing some flexibility on the cost of 4D-Var.

REMARK 4

A variant of $\wp'_{4D}$ is the quadratic problem

$$\wp''_{4D}\text{: minimise } J(\delta x(t_o)) = \tfrac{1}{2}\delta x(t_o)^t\, B^{-1}\delta x(t_o) + \tfrac{1}{2}\sum_{i=o}^{N} (y_{b,i}+H'_i\delta x(t_i)-y_i)^t\, O_i^{-1}\, (y_{b,i}+H'_i\delta x(t_i)-y_i) \qquad (8)$$

with $\quad \delta x(t_i) = \mathfrak{R}(t_i,t_o)\delta x(t_o)$

and $\quad y_{b,i} = H_i[M(t_i,t_o)x_b]$

The cost of $\wp'_{4D}$ and $\wp''_{4D}$ are similar but the storage requirement for the background trajectory is different: in $\wp'_{4D}$ it is the background vertical column at the observation point and in $\wp''_{4D}$ it is the observation equivalent of the background which have to be stored. In $\wp'_{4D}$, $\mathfrak{R}$ is an approximate linearisation of M, similarly in $\wp''_{4D}$, H'_i is an approximate linearisation of H_i.

The structure functions used in the current operational at ECMWF T213 optimal interpolation have a cut-off at wave number 63 (*Lönnberg*, 1988). If we were to use a T106 truncation for $\mathfrak{R}$, this would already be an enhancement in terms of resolution. An adiabatic version for $\mathfrak{R}$ with some basic simplified diabatic processes like horizontal and vertical diffusion and surface friction would produce the same benefits in terms of implicit flow dependent structure functions as obtained by *Thépaut et al.* (1993a).

The CPU cost of an adiabatic semi-Lagrangian T106 L31 model is typically 1/16 of the CPU cost of the T213 L31 version. The gain is then of one order of magnitude.

REMARK 5

Another variant of $\wp'_{4D}$ can be obtained by replacing $H_i'\delta x(t_i)$ in Eq. (8) by a finite difference:

$$\wp'''_{4D} \text{ minimise } J(\delta x(t_o)) = ½\, \delta x(t_o)^t\, B^{-1}\, \delta x(t_o) \tag{9}$$

$$+ ½ \sum_{i=o}^{N} (y_{b,i} - \tilde{y}_{b,i} + \tilde{y}_i - y_i)^t\, O_i^{-1} (y_{b,i} - \tilde{y}_{b,i} + \tilde{y}_i - y_i)^t$$

with $\tilde{y}_{b,i} = H_i\tilde{M}(t_i,t_o)(x_b(t_o))$ being the model equivalent of the observation obtained from a simplified (low resolution, adiabatic) background trajectory and $\tilde{y}_i$ being the model equivalent of the observation obtained from the simplified trajectory issued from $\tilde{x}_b(t_o) + \delta x(t_o)$. $\tilde{x}_b(t_o)$ is the background but at lower resolution. The practical advantage of this formulation over $\wp''_{4D}$ or $\wp'_{4D}$ is that there is less technical development required once the full 4D-Var problem has already been implemented. It is with this formulation that all the numerical experimentations using the incremental approach and presented later have been performed. These three implementations $\wp'_{4D}$, $\wp''_{4D}$ and $\wp'''_{4D}$ are equivalent in the quasi-linear content. They are expected to behave differently in the presence of strong nonlinearities, however we have not seen any arguments as to why one should be superior to the others.

There are two ways of improving the model $\mathfrak{R}$. Firstly one could increase the horizontal resolution: the main drawback here is the cost involved since the CPU follows a power law close to 3. In addition the trajectory storage and thus the IO also follow a cubic law (quadratic at a given time step but the number of time steps increases linearly).

Secondly, it is necessary to take into account the physics. The experiments performed so far (*Thépaut et al.*, 1993a and *Rabier et al.*, 1993) have used only horizontal and vertical diffusion with a simple surface friction. *Rabier* (1992) showed that large-scale condensation is essential in order to get reasonable humidity fields in the upper troposphere. More generally, it is expected that the important feedback loops present in the model M will have to be described to a reasonable accuracy with $\mathfrak{R}$; this is expected to be of particular importance in the tropics. The automatic methods developed at INRIA will assist us in formulating a series of linear models and their adjoints including progressively more effects of the physics (*Rostaing et al.*, 1993).

In terms of cost this will eventually double the CPU cost of 4D-Var (as the cost of the physical parametrizations is about 50% of the cost of the model) but it will immediately double the storage required for the trajectory (and the related IO). Currently, only t values are stored since the dynamics are nonlinear only with respect to these t values and not the $t - \Delta t$. Since, the physics are nonlinear with respect to $t - \Delta t$ values, they too will have to be stored. It should be pointed out, however, that a 2 time-level semi-Lagrangian scheme would not require this extra storage, but the adjoint of the semi-Lagrangian has to be developed with some potential problems for the validity of the tangent linear approximation for too long time steps (*Li et al*, 1993).

The physics is far more nonlinear than the dynamics. As a consequence, the tangent linear approximation is likely to be less valid for the full model than for the adiabatic version. This means that $\wp'_{4D}$, $\wp''_{4D}$ or $\wp'''_{4D}$ are not necessarily a very good approximation of $\wp_{4D}$. A simple way for accounting some of the nonlinearities in the final analysis is to define a sequence $\wp''_{4D}(n)$ of assimilation (similarly $\wp'_{4D}(n)$ or $\wp'''_{4D}(n)$)

$\wp''_{4D}(n)$: minimise $J(\delta x^n(t_o)) =$

$$½\ (\delta x^n(t_o) + x^{n-1}-x_b)^t\, B^{-1}\, (\delta x^n(t_o) + x^{n-1}-x_b) + ½ \sum_{i=o}^{N} (y_i^{n-1}+H_i'\ \delta x^n(t_i)-y_i)^t\, O_i^{-1}(y_i^{n-1}+H_i'\ \delta x^n(t_i)-y_i) \qquad (10)$$

with $\delta x(t_i) = \mathfrak{R}(t_i,t_o)\delta x(t_o)$ (11)

and $y_i^{n-1} = H_i[M(t_i,t_o)\ (x^{n-1}+\delta^{*n-1}x(t_o))]$ (12)

$\delta^{*n-1}x(t_o)$ is the result of the (partial) minimization of $\wp''_{4D}(n-1)$ and $\delta^{*o}x(t_o) = 0$

and $x^{n-1} = x^{n-2} + \delta^{*n-1}x(t_o)$ $\qquad x^o = x_b$, $\ x^{-1} = x_b$

This algorithm can be seen as a pair of nested loops. The outer loop uses the complete model in Eq. (12) to re-define the model trajectory at each iteration of the outer loop. The inner loop uses the tangent linear and adjoint of a simpler (e.g. adiabatic) model (Eq. 11) to minimize the cost function (Eq. 10) for the increments calculated with respect to the re-defined trajectory.

This approach allows a progressive inclusion of physical processes without dealing with large-scale non-differentiable minimization problems, of which little is known in practice. The drawback is that we have no guarantee that the sequence $\delta^{*n}x(t_o)$ will converge. Experimental work is necessary to address this issue but we have to be pragmatic. Highly non regular problems will remain intractable for a long time but this approach is reasonable and probably robust.

REMARK

$\mathfrak{R}$ does not have to be kept constant in this iterative process and one can imagine a sequence $\mathfrak{R}^n$ where the resolution and the number of physical processes dealt with increase with n. It then has a lot of similarities with a multigrid approach.

In this section, we have presented the incremental approach which is the basis for operational implementation of 3D-Var and 4D-Var. It has two advantages:

- It reduces the cost of 4D-Var (also for 3D-Var but this is not a real issue).

- It reduces the memory requirements of 3D-Var (and 4D-Var). Major technical work would be necessary to fit a T213 3D-Var in the 128 M words of the C90 memory.

A careful look at the previous equations shows that the ingredients of the incremental approach are the same as for the original formulation (e.g. 3), namely a background term and an observation term. The only extra requirement is to be able to compare the forecast to the observations (which already existed in 4D-Var) and to store in the observation file the model equivalent of the observation at high resolution $y_{b,i}$ and at low resolution $\tilde{y}_{b,i}$.

3. BACKGROUND TERM

3.1 Univariate formulation

The background term is given by Eq. (13):

$$J_b = \tfrac{1}{2}(x - x_b)^t \, B^{-1}(x - x_b) \tag{13}$$

The practical difficulty is the size of B which does not allow, in practice, its inversion. B, being symmetric, can be diagonalised:

$$B = \mathfrak{L} \, \Lambda \, \mathfrak{L}^{-1}$$

with $\mathfrak{L}$ unitary $\mathfrak{L}^{-1} = \mathfrak{L}^t$

and J_b becomes

$$J_b = \tfrac{1}{2} \, [\Lambda^{-1/2}\mathfrak{L}^{-1}(x-x_b)]^t \, [\Lambda^{-1/2}\mathfrak{L}^{-1}(x-x_b)].$$

The idea of the practical implementation of 3D-Var is to approximate $\mathcal{L}^{-1}$ with a sequence of operators. For a univariate analysis we choose to have the following sequence of operations:

1) Difference x and x_b in spectral space

2) Convert from vorticity, divergence to winds W^{-1}

3) Transform to grid-point space S^{-1}

4) Normalize with respect to background errors σ N

5) Transform to spectral space S

6) Convert from winds to vorticity, divergence W

7) Multiply by the square root of the inverse spectral horizontal background error correlation matrix $h^{-1/2}$

8) Project onto the eigenvectors of the vertical background error correlation matrices P^{-1}

Defining $$\chi = P^{-1}\, h^{-1/2}\, W\, S\, N\, S^{-1}\, W^{-1}\, (x - x_b) = \mathcal{L}^{-1}(x-x_b) \qquad (14)$$

then $$J_b = \tfrac{1}{2}\, \chi^t\, \Lambda^{-1}\, \chi \qquad (15)$$

and $$\nabla_\chi\, J_b = \Lambda^{-1}\, \chi . \qquad (16)$$

Identifying $\mathcal{L}^{-1}$ as the sequence of operators 2) through 8), Λ is a diagonal matrix containing the eigenvalues of the vertical background error correlation matrices.

In order to obtain the gradient with respect to x the adjoints of the above operations have to be applied in reverse sequence.

$$\nabla_x\, J_b = (W^{-1})^*(S^{-1})^* N^* S^* W^* (h^{-1/2})^* (P^{-1})^* \nabla_\chi\, J_b \qquad (17)$$

where * denotes an adjoint operator.

This formulation cannot be expected to describe any operator $\mathcal{L}$ and thus any covariance matrix B. In the 2-dimensional spherical case, it restricts to isotropic correlations but with a full variation of the standard deviation of errors in grid point space. In the three-dimensional case, it allows for non separability.

3.2 Choice of control variable

Let us assume that the minimization is performed in the space of the model variable x. Since the σ's are spatially varying the B matrix will be far from diagonal, and the problem might be ill conditioned if a diagonal matrix is used for defining the scalar product. However, if the σ values are taken as constant over

η levels and errors are assumed uncorrelated in the vertical, then $\boldsymbol{B}$ is diagonal, and exact minimization of J_b alone can be accomplished in 1 iteration using $\boldsymbol{B}^{-1}$ for defining the metric.

Alternatively, the control variable may be taken as χ then the Hessian is simply Λ^{-1}, which is diagonal. Such a change of control variable and the use of the matrix Λ^{-1} for defining the metric improves the preconditioning. The exact solution for J_b alone is found in a single step, even with full geographical variability of the forecast errors and vertical coupling. It is, of course, simpler to re-define the control variable as $\Lambda^{-1/2}\chi$, which reduces J_b to its simplest form - which is done in practice.

3.3 Multivariate analysis

In the variational analysis which became operational in June 1991 at the National Meteorological Center in Washington (*Parrish and Derber*, 1992), the analysis increments are constrained in order to stay close to the equilibrium of the linear balance equation applied on the model levels (slightly modified to account for divergence). This is achieved by a choice of control variable which takes into account the balance equation. The NMC analysis variables are:

- Departures from the 6 h forecast for vorticity and divergence;

- Departures from the balance equation solution of the temperature and surface pressure departures to the 6 h forecast.

By assigning appropriate statistics to the errors of these variables, a balance is achieved which the authors claim is good enough to obviate the need for normal mode initialisation.

In the assimilation context, one is interested in finding an analyzed state which is close both to the observations and to the slow manifold. In the previous sections, the 3D-Var problem was expressed as

$$\min_{x \in E} J(x) = J_b + J_o \tag{18}$$

find the minimum of J in the space E which is the phase space of the model.

Forecast evolution is confined to the attractor of the model, which is approximated by the slow manifold. A consequence is that the forecast errors lie, to first order of approximation, on the tangent plane of this slow manifold. In other words they do not span the whole phase space E but only a subspace, which one may denote E_R. In the current ECMWF operational implementation of OI, E_R is defined by geostrophic balance on the f-plane. In the NMC implementation of 3D-Var (*Parrish and Derber*, 1992), E_R is defined by the linear balance equation $\nabla^2\phi = \nabla.(f\nabla\psi)$ where ϕ is the geopotential and ψ is the stream function, with some enhancements to account for divergence.

A consequence of assuming that the errors are wholly within E_R is that B is singular: the kernel of B contains the orthogonal of E_R which will be denoted by E_G. B is no longer invertible, the formulation defined by (8) and (13) is then no longer suitable for this problem. It can however easily be reformulated as

$$\min_{x\in E_R} J(x) = J_b + J_o \qquad (19)$$

with

$$J_b = [S_R(x-x_b)]^t B_R^{-1}[S_R(x-x_b)] \qquad (20)$$

where S_R is the projection onto the subspace E_R of Rossby modes and parallel to the subspace E_G of gravity modes. The last equation is equivalent to

$$J_b = (x-x_b)^t\, S_R^t\, \tilde{B}^{-1}\, S_R(x-x_b) \qquad (21)$$

where $\tilde{B}^{-1}$ is any matrix identical to B_R^{-1} on E_R and which could take any value on E_G. Furthermore, one should notice that

$$\min_{x\in E_R} J(x) \Leftrightarrow S_R\left[\min_{x\in E} J(S_R(x))\right]$$

and as, in practice, a descent algorithm is used which computes descent directions as a linear combination of several gradients, and as the initial point of the minimization can be assumed to be on the slow manifold, one has the algorithmic equivalence

$$\min_{x\in E_R} J(x) \overset{alg}{\Leftrightarrow} \min_{x\in E} J(S_R(x)) \qquad (22)$$

with

$$J_b = (x-x_b)^t\, S_R^t\, \tilde{B}^{-1}\, S_R(x-x_b) \qquad (23)$$

REMARK

This is different from the problem

$$\min_{x\in E_R} J(x) = J_b + J_o \qquad (24)$$

with

$$J_b = (x-x_b)^t\, \tilde{B}^{-1}(x-x_b) \qquad (25)$$

which could be solved by resetting the gradient in the gravity part to zero. Since however, $S_G(x-x_b) = 0$ at the beginning of the minimization, the two formulations lead to similar results.

3.3.1 *Shallow-water illustration*

From (23) it is clear that if a matrix $\tilde{B}^{-1}$ has been specified and if any kind of independent gravity wave control is used, like in *Courtier and Talagrand* (1990) the inverse of the effective matrix of covariance of first-guess error becomes $S_R^t \tilde{B}^{-1} S_R$. In the following simple example we shall see that, as a consequence, the statistics of the short range forecast errors effectively used are not those which are specified (and there are no simple practical ways of going from one to the other).

Consider the implementation of J_b as it pertains to the shallow-water problem. The state variable of the model $x = (\zeta,D,\phi)$ consists of vorticity ζ, divergence D, and geopotential ϕ. One may define the intermediate variables

$$\tilde{\zeta} = \nabla \times \frac{\vec{v}-\vec{v}_g}{\sigma_{\vec{v}}} \tag{26}$$

$$\tilde{D} = \nabla\cdot\frac{\vec{v}-\vec{v}_g}{\sigma_{\vec{v}}} \tag{27}$$

$$\tilde{\phi} = \frac{\phi-\phi_g}{\sigma_\phi} \tag{28}$$

Isotropy is assumed for the autocorrelation function of $\tilde{\zeta}$, $\tilde{D}$ and $\tilde{\phi}$. The first-guess term J_b expressed in spectral space becomes

$$J_b = \sum_n \left[\frac{1}{a_\phi^n} \sum_m \tilde{\phi}_n^{m2} + \frac{1}{a_D^n} \sum_m \tilde{D}_n^{m2} + \frac{1}{a_\zeta^n} \sum_m \tilde{\zeta}_n^{m2} \right] \tag{29}$$

The a_ϕ^n, a_D^n and a_ζ^n are the expansion of the autocorrelation functions of the fields $\tilde{\zeta}$, $\tilde{D}$, and $\tilde{\phi}$. They can be easily deduced from the grid point values of the autocorrelation function or may be specified directly in spectral space as we shall see in section 3.7.

3.3.2 *A simple example of the impact of imposing a balance*

Consider in the previous example a single wavenumber. For the (n,m) considered, the matrix $\tilde{B}$ is

$$\tilde{B} = \begin{pmatrix} a_\zeta^n & 0 & 0 \\ 0 & a_D^n & 0 \\ 0 & 0 & a_\Phi^n \end{pmatrix} \tag{30}$$

Now assume that the balance condition which is imposed is

$$\begin{cases} \tilde{D}_n^m = 0 \\ \tilde{\Phi}_m^n = \alpha \tilde{\zeta}_m^n \end{cases} \tag{31}$$

where α can be any number. The projection operator S_R is then such that the vector **(0,1,0)** is in the kernel. The vector **(1,0,α)**, being in balance, is kept unchanged and the vector **($-\alpha$,0,1)** is in the kernel. For the latter, it is implicitly assumed that the scalar product is the usual one defined by $(\tilde{\zeta}_m^n)^2 + (\tilde{D}_m^n)^2 + (\tilde{\Phi}_m^n)^2$. Actually there should be some scaling between momentum and mass but it does not change the point being made in this section. The matrix of the projection is then

$$S_R = \frac{1}{1+\alpha^2} \begin{pmatrix} 1 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha & 0 & \alpha^2 \end{pmatrix} \tag{32}$$

The effective covariance matrix of the background errors becomes

$$S_R \tilde{P} S_R^t = \frac{a_\zeta^n + \alpha^2 a_\Phi^n}{(1+\alpha^2)^2} \begin{pmatrix} 1 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha & 0 & \alpha^2 \end{pmatrix} \tag{33}$$

This has a number of implications:

- Assuming that $a_\zeta^n = a_\Phi^n$ and that $\alpha = 1$, which is reasonable for scales close to the Rossby radius of deformation, the effective matrix is half of that which has been specified. In other words, the variance of error which has been specified in grid point space is not that which is actually used, only half of it has an effective contribution.

- Assuming that the specification of the a_Φ^n and a_ζ^n has not been made consistently with the balance equation, then the effective matrix may well depend only on one of the a^n and not on the other for this particular scale.

One could have increments in geostrophic balance by using a univariate background term and controlling independently the gravity waves. The simple example described above shows that such an approach is not an acceptable solution in practice, since one would not be able to deduce from the specified σ_b the value effectively used. It is thus necessary to have the geostrophy embedded within the covariance matrices of the background errors $\underline{B}$.

3.4 Implementation of a multivariate J_b

The basic idea is to split the J_b cost function into Rossby and gravity components and penalise the latter, thus ensuring the analysis increments lie close to the tangent plane of the slow manifold. For the higher vertical modes it does not necessarily make sense to treat the Rossby and gravity parts differently as the frequencies of the latter are no longer so large. These higher vertical modes may be conveniently treated as univariate. The projection on the Rossby and Gravity subspace is achieved using linear initialisation (either explicit or implicit). This may be achieved as follows:

1.1 Difference x and x_b in spectral space ΔX

1.2 Project onto Rossby subspace for desired subset of vertical modes Δx_R

1.3 Project onto Gravity subspace for same subset of vertical modes Δx_G

1.4 Obtain Δx_U by differencing Δx with Δx_R and Δx_G

$$\Delta x_R = R(x-x_b) \tag{34}$$

$$\Delta x_G = G(x-x_b) \tag{35}$$

$$\Delta x_U = \Delta x - \Delta x_R - \Delta x_G \tag{36}$$

It is convenient to split J_b into "slow", "fast" and "univariate" components:

$$\begin{aligned} J_b = {} & \tfrac{1}{2}\, c_R \left(\frac{\Delta x_R}{\sigma_R}\right)^t (h_R^{-\frac{1}{2}})^t\, (V_R^{-\frac{1}{2}})^t\, V_R^{-\frac{1}{2}}\, h_R^{-\frac{1}{2}} \left(\frac{\Delta x_R}{\sigma_R}\right) \\ & + \tfrac{1}{2}\, c_G \left(\frac{\Delta x_G}{\sigma_G}\right)^t (h_G^{-\frac{1}{2}})^t\, (V_G^{-\frac{1}{2}})^t\, V_G^{-\frac{1}{2}}\, h_G^{-\frac{1}{2}} \left(\frac{\Delta x_G}{\sigma_G}\right) \\ & + \tfrac{1}{2} \left(\frac{\Delta x_U}{\sigma_U}\right)^t (h_U^{-\frac{1}{2}})^t\, (V_U^{-\frac{1}{2}})^t\, V_U^{-\frac{1}{2}}\, h_U^{-\frac{1}{2}} \left(\frac{\Delta x}{\sigma_U}\right) \end{aligned} \tag{37}$$

where $(\Delta x/\sigma)$ is a short hand notation for the sequence of operators $(WSNS^{-1}W^{-1})$ as described in section 3.1.

If one defines

$$\chi_R = P_R^{-1}\, h_R^{-1/2}\, WSN_R\, S^{-1} W^{-1} \Delta x_R, \tag{38}$$

$$\chi_G = P_G^{-1}\, h_G^{-1/2}\, WSN_G\, S^{-1} W^{-1} \Delta x_G, \tag{39}$$

$$\chi_U = P_U^{-1}\, h_U^{-1/2}\, WSN_U\, S^{-1} W^{-1} \Delta x_U \tag{40}$$

then, (37) reduces to

$$J_b = 1/2\, c_R\, \chi_R^t\, \Lambda_R^{-1}\, \chi_R + 1/2\, c_G\, \chi_G^t\, \Lambda_G^{-1}\, \chi_G + 1/2\, \chi_U^t\, \Lambda_U^{-1}\, \chi_U$$

For the bulk of the spectrum c_G is set to $\frac{1}{2\varepsilon}$ and c_R to $\frac{1}{2(1-\varepsilon)}$.

ε controls the relative contributions of the first two terms. It may be thought of as the percentage error variance explained by the gravity wave part of the flow. For large-scale gravity modes, for example those important in the description of tides c_G is set equal to c_R, thus these modes will be analyzed with the same weight as Rossby modes.

The Rossby part of J_b is evaluated in vorticity, divergence and mass space and not in Rossby space, otherwise this would have preempted the use of implicit linear initialisation for the separation Rossby-Gravity. Then only the vorticity contribution is included in the cost function since including the divergence and mass would have required the standard deviation of errors in geostrophic balance, similarly the Gravity contribution is evaluated only in mass and divergence space. As we are using mixed implicit-explicit linear initialisation for the separation Rossby-Gravity, a desirable amelioration of the large scale treatment would be to treat the explicit part in Hough mode space and to apply the current formulation only for the implicit part where it is far more valid.

The J_b gradient is now also split into three terms and the adjoint computations 2) → 8) of section 3.1 have to be carried out for each of the terms in turn:-

$$\nabla_{\Delta x_R} J_{b_R} = (W^{-1})^*\, (S^{-1})^*\, N_R^*\, S^*\, W^*\, (h_R^{-1/2})^*\, (P_R^{-1})^*\, c_R\, \chi_R \tag{41}$$

$$\nabla_{\Delta x_G} J_{b_G} = (W^{-1})^*\, (S^{-1})^*\, N_G^*\, S^*\, W^*\, (h_G^{-1/2})^*\, (P_G^{-1})^*\, c_G\, \chi_G \tag{42}$$

$$\nabla_{\Delta x_U} J_{b_U} = (W^{-1})^*\, (S^{-1})^*\, N_U^*\, S^*\, W^*\, (h_U^{-1/2})^*\, (P_U^{-1})^*\, \chi_U \tag{43}$$

the adjoints of 1.1) → 1.4) then gives $\nabla_x J_b$.

The above formulation is fairly general, allowing, in principle, different standard error fields for "fast", "slow" and univariate terms, different horizontal structure functions and different vertical structure functions. For the initial configuration it has been decided to opt for the simplest case of

$$\sigma_U = \sigma_R = \sigma_G, \quad h_U = h_R = h_G, \quad V_U = V_R = V_G$$

which implies $P_U = P_R = P_G$ and $N_U = N_R = N_G$. These constraints can be relaxed as experience dictates.

3.5 Vertical interpolation of fields and effective σ_b

In order to compute observation departures, the model field is interpolated both horizontally and vertically to the observation location. This interpolation can, depending on the location of the observation relative to the model grid and the degree of correlation of background errors, significantly reduce the effective σ_b.

For an observation lying between two model levels with temperature T_1 and T_2 the interpolated model value at the observation point (using linear interpolation) is given by

$$T = \alpha T_1 + (1-\alpha)T_2 \tag{44}$$

and the error in the interpolated value is given by:

$$\varepsilon = \alpha\ \varepsilon_{T_1} + (1-\alpha)\ \varepsilon_{T_2} + \varepsilon_p \tag{45}$$

where ε_p is the error in interpolation process.

Assuming ε_p is uncorrelated with $\varepsilon_{T_1}, \varepsilon_{T_2}$ and $\sigma^2 = < \varepsilon_{T_1}, \varepsilon_{T_1} > = <\varepsilon_{T_2}, \varepsilon_{T_2}>$ and $\beta\sigma^2 = < \varepsilon_{T_1}, \varepsilon_{T_2} >$

and $\sigma_p^2 = < \varepsilon_p, \varepsilon_p >$, with $-1 \le \beta \le 1$,

then $$\sigma_b^2 = (\alpha\ \ 1-\alpha)\begin{pmatrix}1 & \beta\\ \beta & 1\end{pmatrix}\begin{pmatrix}\alpha\\ 1-\alpha\end{pmatrix}\sigma^2 + \sigma_p^2. \tag{46}$$

More generally, for interpolation/extrapolation of one vertical profile $\underset{\sim}{x}_2$ from another $\underset{\sim}{x}_1$ using a linear operator $\underset{\sim}{D}$:

$$\underset{\sim}{x}_2 = \underset{\sim}{D} \cdot \underset{\sim}{x}_1 \tag{47}$$

the error covariance of $\underset{\sim}{x}_2$ is given by

$$\underset{\sim}{C}_2 = \underset{\sim}{D} \cdot \underset{\sim}{C}_1 \cdot \underset{\sim}{D}^T + \underset{\sim}{G} \tag{48}$$

where $\underset{\sim}{G}$ is the error covariance of the operation $\underset{\sim}{D}$. This problem is discussed by *Eyre* (1989) in the context of 1D-Var.

If $\underset{\sim}{G}$ is omitted, then one obtains the (usually false) result that the interpolated profile is <u>more</u> accurate than the profile from which it is interpolated. Examining (46), for $\beta = 0$ one finds that

$$\sigma_b^2 = \alpha^2 \sigma^2 + (1-\alpha)^2 \sigma^2 \tag{49}$$

which for $\alpha = 0$ or $\alpha = 1$ gives $\alpha_b^2 = \sigma^2$

but for $\alpha = 0.5$, gives $\sigma_b^2 = ½\sigma^2$.

This indicates that the background is more accurate at intermediate levels. This is not surprising since an interpolation acts as a filter and there is less variance in the interpolated field. However, it means that the representativeness error is higher and the observation error should be increased accordingly. The analysis increments are controlled directly by the ratio $\dfrac{\sigma_b^2}{\sigma_o^2}$ where σ_o^2 is the observation variance. If σ_b^2 is halved, the observation will have less impact on the analysis. For realistic values of σ_o and σ_b, the analysis increments are smaller by 20%. This is important during the validation of J_b since the analysis increments depend significantly on their vertical location.

For $\beta = 1$ one obtains $\sigma_b^2 = \sigma^2$

and for $\beta = -1$ one obtains $\sigma_b^2 = \sigma^2(1-2\alpha)^2$.

This analysis indicates that an observation lying midway between model levels is given less and less weight as the structure functions become increasingly sharp. Ultimately, as the correlation between adjacent levels becomes negative the observation is ignored. This is a genuine problem since situations exist where an observation is of no use. It is also a design feature of 3D-Var. It reveals a problem which becomes acute when the structure functions are too sharp compared to the model vertical discretization.

The feature does not show up in the current ECMWF implementation of OI. However, in the ECMWF OI, the background is interpolated to the observation point using (44) but the σ's which have to be explicitly interpolated are not interpolated using (46) but using (44). OI is thus not mathematically consistent (in its implementation!)

One could minimize the impact of the problem by using cubic interpolation in the vertical. One could also use tricks to mimic OI such as explicitly changing σ_o so as to compensate for the changed σ_b. However, it is not clear if one should do this. An observation located at a half level is probably not as informative as an observation located at a full level.

It is clear that one should ensure that the vertical structure functions are reasonably resolved by the vertical discretization of the model and, if at some levels they remain too sharp (e.g. at the tropopause), it is a strong argument for having more vertical resolution in the model at such levels. For multilevel observations, one should have an observation operator (and its associated error covariance) which correctly represent the observation process and any inherent vertical averaging. In practice for using part B and D of temperature messages, one could extract data from the continuous vertical profile such that their vertical resolution is consistent with the model vertical resolution. It is worth noting that the problem also occurs in the horizontal but is of smaller magnitude since i) we use bicubic interpolation, and ii) the horizontal correlation of errors are relatively broad compared to the horizontal resolution.

3.6 Transformations between P and T, lnp_s

The model uses a linearised mass variable $P=\phi + R_d\overline{T}\ \ln p_s$, where ϕ is linearised geopotential height and $\overline{T}$ is a (constant) reference temperature. As the multivariate formulation uses the model's Hough modes to distinguish between balanced and unbalanced components of the fields it is necessary to work in terms of P rather than temperature and log surface pressure.

However, the advantage of the vertical coupling is not gained without some loss. There are $(L+1)$ degrees of freedom in the T, $\ln p_s$ combination, but only L in P, where L is the number of model levels. Clearly there is no problem in defining P from T and $\ln p_s$, but the transformation from P to T and $\ln p_s$ involves a degree of arbitrariness. It is interesting to study this latter transform in more detail. The transform from T and $\ln p_s$ to P may be expressed as $P = G\ V$ where G is a L by $(L+1)$ matrix and V is a vector of dimension $(L+1)$ containing L temperature values plus log surface pressure, P is a vector of dimension L. G basically contains the linearised hydrostatic integral of temperature to obtain geopotential height. In a similar way, the inverse transform may be expressed as $V = H\ P$, where H, the pseudo inverse of G for the energy matrix, is a $(L+1)$ by L matrix. The transform from T, $\ln p_s$ to P and back to T, $\ln p_s$ may be written $V' = E\ V$ where $E = H\ G$, and E is of dimension $(L+1)$ by $(L+1)$. If one calculates the eigenvalues and eigenvectors of the matrix E one finds L eigenvalues equal to 1 and one eigenvalue equal to zero. The latter eigenvalue is associated with the kernel, or "nullspace", of the matrix E. The structure of the kernel, at grid scale in the vertical, describes the information which is lost in the transformation from T, $\ln p_s$ to P. This information cannot be reinstated by the transform from P to T, $\ln p_s$ and the final field is characterised by a zero projection onto the kernel. As an example, if one takes a temperature profile, and applies the operator E, one obtains temperature differences between the two profiles at model level 19 as high as 159° C!

Clearly, this is not a very good description of the original temperature profile. The information lost is vital to a correct description of the profile.

The problem is how to deduce the correct amplitude of the kernel. In the above example there is a clear solution: one calculates the amplitude of the projection of the original field onto the kernel, applies the operator E, then simply adds back the kernel with its original amplitude. This technique reproduces the original profile to within machine accuracy.

If one does not know the original T, $\ln(p_s)$ field (which is the case with variational analysis since the control variable is a function of P) then the amplitude of the kernel has to be diagnosed in some way. In order to do so it is necessary to close the problem by applying an additional constraint. An obvious choice is one which minimizes the second derivative of temperature in the vertical. Define a matrix S such that

$$S = \begin{pmatrix} 1+a_2 & -2 & 1-a_2 & 0 & . & . \\ 0 & 1+a_3 & -2 & 1-a_3 & . & . \\ 0 & 0 & 1+a_4 & -2 & . & . \\ 0 & 0 & 0 & 1+a_5 & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 1+a_{n-1} & -2 & 1-a_{n-1} \end{pmatrix}$$

where S is of dimension L by $(L-2)$. The 'a' terms are introduced by the irregular spacing of the eta-levels. A measure of the 'noise' in the profile is then given by $J = T'^t S^t S T'$, where T' is the temperature profile as derived from P. What is the amplitude of the kernel such as to minimize J? Define a vector L containing the first L elements of the kernel K (the one dropped is that operating on surface pressure), then the problem may then be expressed:-

find the value c such that $\frac{\partial}{\partial c}\,(T' - cL)^t\, S^t\, S\, (T' - cL) = 0$. This is a linear equation in c, the root of which is easily determined. A little algebra leads to the result that $c = \frac{T'^t\, S^t S\, L}{L^t\, S^t S\, L}$; note that the denominator is a constant. One may precalculate a vector $Z = (S^t S\, L)\, /\, (L^t\, S^t S\, L)$ then $c = T^t\, Z$.

Carrying this out (i.e. adding cK to the T, $\ln p_s$ field derived from P) gives a profile. At most levels the difference with the original profile is about tenth of a degree, the largest difference, of about a half degree, occurs at the lowest level. It is important to note that this approach only changes the amplitude of the kernel. There is no change to the corresponding P field. Information on P supplied through the minimization is passed on intact.

EFFECT ON VERY SHARP PROFILES

The effect of the transformation on a delta function at model level 11 was studied. In this case the field as recovered from P shows rather a lot of noise, the spurious signal at level 19 is 30% of the value at level 11. It should be noted that the criterion for diagnosing the amplitude of the kernel (minimum second derivative of T in the vertical) is clearly inappropriate when trying to recover a delta function. Even so the technique reduces this noise by over a factor of two. This is a problem in 3D-Var because of the sharpness of the temperature structure functions.

NMC APPROACH

According to *Parrish and Derber* (1992), the NMC approach is to explicitly minimize a cost function for temperature $J = T^t\, S^t\, S\, T$ where the matrix S is an $(L-2) \times L$ matrix with all zeros except for the three diagonals $S_{jj} = 1$, $S_{jj+1} = -2$, $S_{j,j_2} = 1$, $1 \le j \le L - 2$. Which, for regularly spaced levels, applies 2nd derivatives to the temperature in the vertical. First, the equation for J is expressed in terms of P and $\ln p_s$ using the definition of P. The resulting equation is then minimized, at constant P, to solve for $\ln p_s$. Finally P and this value of $\ln p_s$ is used to solve for T using the definition of P. The result should be very similar to that obtained using our approach. Derber (personal communication) told us that the noise in assimilation can be as high as 4K in the stratosphere (where no physical process can stop it growing).

ANOTHER APPROACH

Two weaknesses of the previous approaches are firstly that they are not local: a global criterion is minimized and secondly that they make no use of the vertical structure functions. Let us denote by $\boldsymbol{B}_{T,p_s}$ the vertical covariance matrix in the $\boldsymbol{T,p_s}$ space and by $\boldsymbol{B_P}$ in $\boldsymbol{P}$ space. $\boldsymbol{B_P}$ and $\boldsymbol{B}_{T,p_s}$ are related by

$$\boldsymbol{B_P} = \boldsymbol{G}\, \boldsymbol{B}_{T,P_s}\, \boldsymbol{G}^t.$$

A statistically optimal solution is to introduce the pseudo inverse $\boldsymbol{C}$ of $\boldsymbol{G}$ but for the two spaces with respective metric $\boldsymbol{B_P}$ and $\boldsymbol{B}_{T,P_s}$. The expression for $\boldsymbol{C}$ is then

$$\boldsymbol{C} = \boldsymbol{B}_{T,P_s}\, \boldsymbol{G}^t\, \boldsymbol{B}_P^{-1} \tag{50}$$

and works in practice in that the effective covariance matrix $\boldsymbol{C}\, \boldsymbol{B_P}\, \boldsymbol{C}^t$ remains close to $\boldsymbol{B}_{T,p_s}$ within one percent. What this approach requires for being successful is that there is little statistical structure lost while going from covariances specified in $\boldsymbol{T,p_s}$ space to covariances specified in $\boldsymbol{P}$ space. This approach is far more successful than the other two in controlling the 2-grid noise. In our implementation, as J_b is computed in terms of vorticity and the noise controlled in terms of the mass variable, it is critical to have the vertical profile of vorticity standard deviations of errors proportional to the vertical profile of $\boldsymbol{P}$ standard deviation of errors.

3.7 Horizontal structure functions

3.7.1 *Observational evidence*

Published literature concerning the shape of horizontal structure functions of short-range forecast errors is somewhat scarce. The information at our disposal consists of the two papers *Hollingsworth and Lönnberg* (1986) (hereafter HL86) and *Lönnberg and Hollingsworth* (1986) (hereafter LH86) which were the basis for the structure functions of the "new" ECMWF OI described by *Shaw et al.* (1987). *Lönnberg* (1988) (hereafter L88) describes some "revised structure functions" as used at ECMWF. From elsewhere, we have *Bartello and Mitchell* (1992) (hereafter BM92) and *Mitchell et al.* (1990). The latter has not been used since no spectra were presented in it.

These authors have been using the northern American radiosonde network (the only homogeneous network available) and thus have sampled scales from 300 km to 3000 km which corresponds to total wave numbers 6 to 66. Comparing Fig. 2 of HL86 with Fig. 11 of BM92, one sees that the spectra agree pretty well in the range 15 to 60 with, in particular, a maximum around wave number 20. The maximum at wave number 9 in HL86 is suspect and this can be explained by an accumulation of larger scale energy which was not properly sampled.

In terms of slope of the wind modal spectrum, BM92 are very careful, saying that it is negative in the range 3-6. However, for geopotential they claim a range of minus 3-4. These figures are contradictory since, under the geostrophic assumption, a slope $-p$ for wind leads to a slope $-(p+2)$ for geopotential. LH86 came to the conclusion of a negative slope for wind modal spectrum of between ½ and 1 which is again contradictory with Fig. 2 of HL86 where the wind slope is in the range minus 3 to 4 and closer to 3 than to 4. The explanation of this apparent paradox is in LH86 Fig. 8 where we can see that the end of the modal spectrum is noisy to the point that it is difficult to infer any sensible geopotential slope in the inertial range. This has been recognised by L88 since the revised structure functions have been obtained using a slope 4 for geopotential modal spectrum.

Fig. 2 of HL86 is believable for the inertial range since:

- it is stable with respect to number of Bessel functions retained
- it is confirmed by Fig. 14 of HL86
- it is confirmed for wave number 15 to 40 by Fig. 8 of LH86.

It has been chosen to use a slope of -2 in the inertial range for the wind power spectrum (which corresponds to -3 in terms of modal spectrum) since it is consistent with *Charney* (1971) theoretical analysis of 2D quasigeostrophic flow.

Fig. 2 of HL86 seems doubtful for the larger scales as wind pairs lead to less information on the large scales than height pairs. Furthermore, it is not in agreement with either Fig. 8 of LH86 or Fig. 11 of BM92. The fact that the slope of height is negative in Fig. 7 of LH86 or Fig. 5 of BM92 and that it is positive in wind (Fig. 8 of LH86 or Fig. 11 of BM92) indicates that the wind slope should be between 0 and 2 assuming geostrophy. For a slope of 0 the wind spectrum is flat and for a slope of 2 it is the height spectrum which would be flat. Using the 2 points available in LH86 lead to a negative slope 1 for height. In BM92 the 2 points would lead to a negative slope 0.3 for height. For wind this transforms to a positive slope in the range 1 to 1.7 (assuming geostrophic balance).

3.7.2 *A parametric formula for the spectrum*

Consider the following expression for the wind power spectrum

$$f(n) = \frac{\varepsilon + \left(\dfrac{n}{n_1}\right)^{p_1}}{1 + \left(\dfrac{n}{n_o}\right)^{p_o+p_1}}$$

with $p_o = 2$

$p_1 = 3$

$n_o = 15$

$n_1 = 2$

$\varepsilon = 0.1$

For large n, $f(n) \sim n^{-p_o}$ which justifies the choice $p_o = 2$.

For small n (and ε) $f(n) \sim n^{p_1}$. Thus p_o gives the (negative) slope in the inertial range and p_1 the (positive) slope for the large scales.

The maximum of $f(n)$ is for $n \geq n_o$ but close to n_o. For given slopes p_o and p_1 the correlation length scale is quite sensitive to the choice of n_o, for the slopes chosen $n_o = 15$ gives a length scale of about 500 km for geopotential.

ε has been made negligible $\varepsilon = 0.1$, it controls the shape (and not slope) of the spectrum for very large scales $n = 0$ to n_1. Once ε is negligible, n_1 is just a scaling factor, it plays then no role in the shape of f. The value chosen for p_1 is based on forecast error studies using satellite radiances.

The spectrum is rescaled so that the correlation for zero separation is equal to 1. The scaling factor is easily computed as

$$\left(\sum_n f(n)\, p_n^o(o)\right)^{-1} = \left(\sum_n f(n) \times \sqrt{2n+1}\right)^{-1}$$

Fig. 1 presents the spectrum obtained and Fig. 2 presents the corresponding <φ,φ> grid point correlation function superimposed on what is currently used in ECMWF OI.

No information was available in this literature on the planetary scales. ε and n_1 are the free parameters of the formulation which have been evaluated in the following sections.

3.7.3 *Large-scale component of the background error correlation estimated from satellite data*

To objectively determine the large-scale contribution of background error correlation using observed increments, global data coverage is needed. Satellite data is the most appropriate data source to be used for this purpose. The radiance measured in a particular satellite channel is effectively a weighted average of the temperature profile T(z)

$$R = \int_z B(T(z)) \frac{d\tau(z)}{dz} dz$$

Thus calculating radiances from the model and comparing to measured values allows an "observation" of model error in this vertically averaged space. Departure fields in brightness temperature from PRESAT output files are used. A departure field is written as $\psi_b - \psi_o$, where ψ_b is the background field and ψ_o the observation. The basic assumption is that the normalized departure $D_i = \dfrac{\psi_b - \psi_o - \overline{(\psi_b - \psi_o)}}{\sigma_{\psi_b - \psi_o}}$ correlations reflect the spatial correlation of the forecast error $(\psi_b - \psi_t)$. This should be true, at least in the large scale, where observation error correlations between the radiances themselves may be ignored. Clear column radiance data from NOAA11 and NOAA12 over both land and sea are used. These data are bias-corrected using an air -mass dependent procedure to remove the bias between model and observation (*Eyre*, 1992). This should further remove any chance of radiance correlations contaminating the results (remark: the bias correction is applied separately for each satellite).

Statistics are computed for two different periods (October 1992 and February 1989). Separate departure fields for 0,6,12,18 UTC are interpolated using a three pass Cressman analysis to produce global fields on a grid of 160 latitudes and 320 longitudes (radii of 5,3 and 1 grid points around each observation are used in the analysis, but the result is not sensitive to this choice for large scale features). The mean $\overline{\psi_b - \psi_o}$ and standard deviation $^{\sigma}\psi_b - \psi_o$ are computed at each grid-point for each of the four 6-hour slots of the 44 (October 92) and 48 (February 89) departure fields which could be successfully analyzed (i.e. had good data coverage). The departure fields are then normalized by removing the mean and dividing by the standard-deviation.

For each realisation i, the normalized departure field $(Di(\lambda,\mu))\lambda,\mu$ is then transformed in spectral space $(Di(n,m))_{n,m}$.

Results in terms of slopes of the auto-correlation function agree to within 10% for the two periods, except for those derived using MSU 4 which gives the noisiest results. Results for February 1989 were generally smoother, as the limb correction was applied to the satellite data for this data set whereas it was not done for the October 1992 data. Results for February 1989 and for 6 channels are presented. Fig. 3a shows the spectrum derived from MSU4 (peaking at 70-100 hPa), 3b HIRS4 (peaking at 400 hPa), 3c MSU2 (peaking at 600-700 hPa), 3d HIRS15 (peaking at 700-800 hPa), 3e HIRS13 (peaking at the ground) and 3f for HIRS 11 (humidity channel peaking at 700 hPa).

The corresponding large-scale slopes ($n = 2$ to **10**, typically) are 0 for MSU4, 0.4 for HIRS4, 0.5 for MSU2, 0.5 for HIRS15, 0.6 for HIRS13, and 0.6 for HIRS11. The stratospheric channel MSU4 exhibits a rather flat spectrum at large-scale (in fact, the spectrum is hardly regular enough to give a significant slope different from 0). In the tropospheric channels, slopes increase from 0.4 in the upper troposphere to 0.6 in the lower troposphere.

3.7.4 *Horizontal correlations derived from NMC's method*

In order to specify their statistics of forecast error in their 3D-Var implementation, NMC accumulates statistics of the differences between forecasts at different ranges valid at the same time (*Parrish and Derber*, 1992). For this study the method is applied to comparisons of 24 and 48 hour forecast fields valid at 12 UTC on the same day. This choice is somewhat arbitrary, but using the 24 hour forecast as a validation avoids any problems associated with spin-up. It also allows differences to be examined over 24 hours, a period not too long in order for the error not to be significantly different from a 6 hour forecast error, and not too short that the forecasts are too similar because of lack of update by the data.

The truncation selected for the computations was T106 L31 and operational forecasts were compared for December 1992. As a first step, the covariance computation was performed directly in spectral space. The mean error field is a full spectral field (locally varying mean) which is removed from each departure field. However, for the necessary division by the standard-deviations to go to correlation matrices, the variances were globally averaged at each level. This is of course a slight approximation, but is believed to be sufficient.

Horizontal correlations were examined for each vertical level. Fig. 4 shows the variance spectra for wind (4a) and temperature (4b) at levels 1 (dotted line), 18 (dashed line) and 31 (solid line). It can be seen from the slopes that in the small-scales much more energy is present in model error close to the ground than high in the stratosphere. In the large scales the differences between the slopes are more subtle. However, for the temperature spectra one can notice a slightly steeper slope for lower levels, which is in agreement with the results obtained from the radiances in section 3.7.3. In Figs. 5 and 6, the autocorrelation spectra for all levels are presented in terms of wind and temperature respectively, for levels 1 to 10 (Fig. 5a and 6a), levels 11 to 20 (Fig. 5b and 6b), levers 21-30 (Fig. 5c and 6c). The top levels show more dispersion, particularly in the small-scale (where there is less energy at the very top of the model). Most of the variation in the shape of the spectrum and then of the length scale is concentrated in the stratosphere as previous studies have already shown (HL86 and LH86).

Three representative levels have been chosen: levels 6 (around 100 hPa), 16 (around 400 hPa) and 26 (around 850 hPa). Autocorrelation spectra are presented in Fig. 7, where level 6 is represented by a dotted line, level 16 by a dashed line and level 26 by a solid line. For each of these spectra, we computed $n_{\mathbf{max}}$ which is the wavenumber where the spectrum reaches its maximum, the length-scale in km, (defined as the component length-scale of HL86) and the slopes in the large-scale ($n = 2$ to 10), in the range ($n = 40$, $n = 70$), and in the range ($n = 70$, $n = 100$). Results are presented in the following table:

	max		Length-scale		Large-scale slope		Slope 40-70		Slope 70-100	
Level 6 (0-24) KE and T	10	8	288	292	0.91	0.35	-3.1	-2.3	-5.7	-4.9
Level 16 (0-24) KE and T	15	10	239	289	1.43	0.78	-2.4	-2.1	-3.4	-2.9
Level 26 (0-24) KE and T	16	12	241	262	1.37	0.85	-2.5	-1.7	-2.8	-2.5
Level 6 (24-48) KE and T	9	10	299	299	1.38	0.64	-2.9	-2.6	-6.0	-5.2
Level 16 (24-48) KE and T	14	10	238	275	2.24	0.91	-2.2	-2.1	-3.9	-3.6
Level 26 (24-48) KE and T	12	10	232	236	1.80	0.94	-2.0	-1.9	-3.3	-3.1

For both the 0-24 and 24-48 model differences, one can notice the predominance of large-scales in the stratosphere, with length-scales bigger than in the troposphere. A rather flat slope is seen in the very large scales ($n = 2$ to 10) and a very steep slope towards the small scales. For the slopes of the temperature spectra in the large scale one notices an increase when going from higher to lower levels. The slopes in the troposphere are typically around 0.8 for the 0-24 differences and 0.9 for the 24-48 differences. The results obtained from the satellite data are around 0.5. Although these results do not give exactly the same figure, they appear to be reasonably consistent. For the kinetic energy slope in the range 40-70 values are between -2 and -3 and in the range 70-100 between -3 and -6. These slopes agree well with the latest results of *Bartello and Mitchell* (1992) who found between -3 and -6 in terms of modal S_n spectrum which corresponds to -2 to -5 in terms of the power P_n spectrum used in this study. The broader correlations at higher levels, which can be seen from the bigger length scales are in agreement with previous observational results (HL86, Fig. 13; LH86, Fig. 6).

We have chosen to continue the study with the 24-48 differences. Although the results are not very different from the 0-24 differences, they are less likely to be contaminated with spin-up problems (it is what is currently used at NMC).

For the three representative levels chosen earlier (6,16 and 26), wind spectra were computed separately for the vorticity part and the divergence part. The divergence errors are in smaller scales than the vorticity errors. The length scale is 191 km as compared to 245 for the rotational part of the wind at level 16.

To check the degree of balance in the model errors derived with this method we also computed the 500 hPa variance spectra for wind and geopotential. These are presented in Fig. 8 (a and b). The length-scales are respectively 239 and 493 km. The value of 493 km for the component length scale of the height auto-correlation agree well with what is currently used in OI (500 km in the latitude band 90N-30N) and correspond to what was initially chosen for 3DVAR.

The slopes for the wind and the geopotential spectra are respectively 1.9 and 0.9 for the range $n = 2$ to 10, -2.1 and -4.2 for the range $n = 40$ to 70, -3.8 and -5.2 for the range $n = 70$ to 100. The value of -4 for the geopotential in the range $n = 40$ to 70 is slightly steeper than what is actually used in OI (equivalent slope of -3, *Lönnberg*, 1988), but corresponds well to results in *Bartello and Mitchell* (1992). Furthermore, it is consistent with a geostrophic assumption, as there is a difference of -2 between the wind and the geopotential spectra slopes. This difference of -2 in the slopes does not apply exactly at smaller scales, or at all at large scales (1.8 and 0.9). Fig. 9 (a and b) shows the geopotential variance spectra which would be obtained by applying both a geostrophic balance ($\phi = f\psi$) and a linear balance ($\Delta\phi = (\nabla . f)\nabla\psi$) to the wind spectra. All the geopotential spectra (i.e. observed, geostrophically balanced, linearly balanced) agree well in the range $n = 10$ to 100, but both balanced spectra disagree with the observed one in the range $n = 0$ to 10. Although the linearly balanced spectrum is in better agreement with the observed one than the simple geostrophically balanced spectrum, it is clear that we should apply a more elaborate relationship between balanced mass and wind increments at these large scales. On the whole, there is 10% less energy in the error variance for the balanced geopotential than for the observed one.

The results obtained using the NMC approach are compatible with the studies of HL86 and *Bartello and Mitchell* (1992) for the synoptic scales. They are also compatible with satellite data for the very large-scale. We have thus chosen to specify the 3D-Var structure functions (horizontal as we have just seen and vertical as we shall see in the next paragraph) from these statistics as it gives access to fields not well resolved by data like e.g. divergence. Fig. 10 presents the spectrum as defined in section 3.7.2 (for wind) superimposed with what is obtained from NMC's approach at 500 mb. The agreement is very good for the synoptic scales and is certainly better than other approximations we are making like the isotropy of the autocorrelation function. For the large scale, we have chosen to be in agreement with the geopotential rather than the wind. The parametric formulation has been used in all the experimentation using separable structure functions. For non separable structure functions, the spectrum displayed in Fig. 5 has been used directly.

3.8 The vertical correlations

In the previous section, the spectra of the variances at each level i were given as

$$\sigma_i^2 = \sum_{n=0}^{N} \sum_{m=-n}^{n} \overline{\psi_n^m(Zi)\, \psi_n^m(Zi)^*}.$$

The covariance between levels Zi and Zj may be written

$$\mathrm{Cov}(Zi, Zj) = \frac{1}{4\pi} \int_\lambda \int_\mu \overline{\psi(\lambda,\mu,Zi)\, \psi(\lambda,\mu,Zj)} d\lambda d\mu$$

$$\mathrm{Cov}(Zi, Zj) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \overline{\psi_n^m(Zi)\, \psi_n^m(Zj)^*}$$

and the corresponding correlations, dividing by globally averaged standard deviations are

$$C(Zi, Zj) = \frac{1}{\sigma_i\, \sigma_j} \mathrm{Cov}(Zi, Zj).$$

The standard-deviations for height are compared with those from LH86 (column 1 below). The height statistics are computed directly from pressure level geopotential fields (ϕ^2) (column 2).

	$\sigma(\phi^2 obs)$	$\sigma(\phi^2)$
10 hPa	50 m	40 m
100 hPa	16 m	16 m
150 hPa	17 m	18 m
200 hPa	17 m	22 m
250 hPa	20 m	25 m
300 hPa	20 m	26 m
400 hPa	16 m	23 m
500 hPa	13 m	19 m
700 hPa	10 m	15 m
850 hPa	10 m	14 m
1000 hPa	10 m	16 m

(NB. the 10 hPa value for $\sigma(\phi^2 obs)$ is the one used in OI at mid-latitudes)

It can be seen that the computed values are slightly larger than those observed in HL86. But they follow similar vertical variations. A maximum at the top (10 hPa) is observed and a minimum at the lower levels. There is also a local maximum at 300 hPa and a local minimum at 100 hPa. Similarly, standard deviations for wind and temperature have been derived: they compare well with HL86/LH86.

Fig. 11 represents the vertical correlations for temperature (11a), rotational part of the wind (11b), divergent part of the wind (11c) and geopotential (11d).

The temperature results agree well with the vertical correlations derived from background-radiosonde statistics shown in Fig. 12, top left panel (currently used in 1D-VAR) and with those from HL86 (derived from thicknesses) reproduced for convenience in Fig. 12, top right panel.

The correlations of the rotational part of the wind and of the geopotential look reasonable, although are slightly broader than in Fig. 16 of HL86 and Fig. 10 of LH86. This can be explained by the fact that HL86 and LH86 statistics are computed over North America. The vertical structure functions used in the operational OI are e.g. broader over the oceans. The vertical correlations of the divergent part of the wind are quite sharp, but these are known to be sharper than the ones for the rotational part (HL86).

The only inconsistency found is between the temperature results and the vertical temperature correlations of the OI over land. There is less discrepancy with the OI vertical correlations over the ocean even if they are still too sharp. These are derived by differentiation of functionally fitted height statistics which have been artificially sharpened for the analysis of wind measurements (Fig. 12, bottom panel). These are much sharper than our results and also exhibit negative side-lobes close to the main peak, a direct consequence of differentiating rather sharp height statistics. Attempts have been made to explain this inconsistency.

Taking the wavenumber range $n = 10$ to $n = 100$ ("synoptic-scale"), vertical correlations were computed from the 48-24 hour forecast statistics. In Fig. 13 results for the rotational part of the wind (13a), the stream function (13b) and the geopotential (13c) are presented. If the geostrophic assumption is valid ($\psi = \phi/f$), the same statistics for streamfunction and geopotential should be observed. Comparing Figs. 13b and 13c, streamfunction correlation structure is seen to be slightly sharper than that for geopotential. This suggests there is some imbalance between wind and geopotential in the forecast error statistics.

Also implicit in the derivation of temperature correlations from the OI statistics is the assumption of separability (i.e. same vertical correlation for each horizontal scale). To investigate the separability assumption, one can introduce a vertical correlation matrix defined separately for each term of the horizontal expansion as in HL86. In our notation, the covariance between two points separated by a distance r on the horizontal at two different horizontal levels Zi and Zj can be written as

$$\overline{\psi(x,Zi)\,\psi(x+r,Zj)} = \sum_{n=o}^{N} \left(P_n(Zi)\;P_n(Zj)\right)^{\frac{1}{2}}\; C_n(Zi,Zj)\;\frac{P_n^o(r)}{\sqrt{2n+1}}$$

where $P_n(Zi) = \sum_{m=-n}^{n} \overline{\psi_n^m(Zi)\;\psi_n^m(Zi)^*}$

$$P_n(Zj) = \sum_{m=-n}^{n} \overline{\psi_n^m(Zj)\;\psi_n^m(Zj)^*}$$

and $C_n(Zi,Zj) = \dfrac{\sum_{m=-n}^{n} \overline{\psi_n^m(Zi)\;\psi_n^m(Zj)^*}}{\left(P_n(Zi)P_n(Zj)\right)^{\frac{1}{2}}}$

If the vertical correlation matrices $C_n(Zi,Zj)$ are independent of n and equal to a matrix $C(Zi,Zj)$, the expansion becomes $\overline{\psi(x,Zi)\,\psi(x+r,Zj)} = C(Zi,Zj) \sum_{n=o}^{N} \left(P_n(Zi)P_n(Zj)\right)^{\frac{1}{2}} \dfrac{P_n^o(r)}{\sqrt{2n+1}}$.

Furthermore, if the horizontal spectra $P_n(Zi)$ are independent of level and equal to P_n, the expression is fully separable $\overline{\psi(x,Zi)\,\psi(x+r,Zj)} = C(Zi,Zj) \sum_{n=o}^{N} P_n \frac{P_n^o(r)}{\sqrt{2n+1}}$. The variation of the vertical correlation matrices $C_n(Zi,Zj)$ with respect to n can be investigated.

In Fig. 14 the vertical correlation matrices of the rotational part of the wind are shown averaged over the wave number ranges $n = 0$ to 10, $n = 10$ to 20, $n = 20$ to 30, $n = 30$ to 40, $n = 40$ to 50 and $n = 50$ to 60. The sharpening of the structure when going to smaller scales is clear for both variables, which questions the validity of the separability hypothesis. This is consistent with the findings in HL86 and LH86, although the dispersion was less marked. A lack of separability is also seen in the vertical correlations for the rotational part of the wind and for the streamfunction being different. Thus, even assuming a perfect geostrophic balance, vertical correlation structure summed over all wavenumbers for wind is sharper than that for geopotential.

The main (and easy to implement) part of the non separability is the dependency of the T, p_s correlation matrix with the total wave number n. As structures become sharper in the vertical for smaller horizontal scales, this can be seen as a 3D-isotropy.

Using these T, p_s correlations and the parametric spectrum as defined in the previous section (not variable with level) one gets the wind vertical correlations displayed in Fig. 15 (dotted line). They agree remarkably well with the vertical structure obtained directly from wind (solid line). A further improvement easy to implement in the 3D-Var formulation is different vertical correlations for divergence as for vorticity; however it has not yet been tested, even if it would sharpen the wind vertical correlations.

4. THE OBSERVATIONS USED

4.1 Conventional observations

By conventional observations we mean ground based observations whose observation errors are not correlated in the horizontal. In addition, the space based satellite wind measurements (SATOB) are included in this group. The following is the list of the observing systems:

- SYNOP: surface pressure (p_s), 10m wind components (u_{10m},v_{10m}), 2m temperature (T_{2m}) and 2m relative humidity (RH_{2m})
- AIREP: upper air wind components (u,v) and temperature (T)
- SATOB: upper air wind components (u,v)

- DRIBU: surface pressure (p_s), 10m wind components (u_{10m},v_{10m}) and 2m temperature (T_{2m})
- TEMP: 10m wind components (u_{10m},v_{10m}), 2m temperature (T_{2m}), 2m relative humidity (RH_{2m}), geopotential heights (Z), upper air wind components (u,v), upper air relative humidity (RH) and upper air temperature (T)
- PILOT: upper air wind components (u,v)

Most of the observed quantities listed above are used by the current ECMWF OI system with the exception of near surface observations. Specifically in the OI, 10m wind and 2m relative humidity observations are used only over sea, whereas 2m temperature is not used at all (10m wind is also used in tropical land areas with low altitudes). Although not all of these quantities are directly observed, as for example relative humidity, we will continue to treat them as observed in order to stay, at this stage, as close as possible to the OI system.

The observation term J_o (2nd term of Eq. 3) is a function of the covariance matrix O of the observation errors and the observation departures from the model ($y-Hx$). Specification of the matrix O (or its inverse) requires a specification of the observation errors; specification of the observation departure requires a specification of the operator H.

It is reasonable to assume that observation errors associated with different observation types are uncorrelated. This assumption is used in most operational analysis systems. Thus following *Pailleux et al.* (1991), O is a block-diagonal matrix and hence J_o can be split in a number of independent terms for each individual observation type:

$$J_o = J_{o(synop)} + J_{o(airep)} + J_{o(satob)} + \dots \tag{51}$$

Within the individual observation type, it is reasonable to assume that observation errors of different observation points are not correlated. Then the contributions from different observations points can be computed independently. For example, the radiosondes' contribution to the cost function can be split:

$$J_{o(temp)} = J_{o(temp)1} + J_{o(temp)2} + \tag{52}$$

Furthermore, an additional assumption for each individual observation can be made: the observation errors of the different observed quantities are uncorrelated, allowing to split the above term even further,

$$J_{o(temp)i} = J_{o(u,v)i} + J_{o(z)i} + J_{o(RH)i} + \tag{53}$$

where each of the terms on the right hand side of (6) is a quadratic form built on the vertical covariance matrix of the observation error. We further assume all observation errors to be uncorrelated in the vertical except for the geopotential height from TEMP observations, where we assume correlation consistent with temperature errors uncorrelated in the vertical.

The observation errors decorrelation assumptions made so far are reasonable for the conventional observation types. However, since the observation error contains the representativeness error one can expect the observation error of adjacent SYNOP temperatures to be correlated in certain weather conditions such a winter inversions over cold land. It should also be borne in mind that the representativeness error includes not only the local (sub-grid) variability, but also the error in the operator $\boldsymbol{H}$. No explicit provisions have yet been made in the ECMWF 3D variational analysis to take account of the observation error correlations induced by the representativeness errors.

In computing the cost function J_o, one of the first steps is to find the observation departure (y-$\boldsymbol{Hx}$). Since, with rare exceptions, no model variable is directly observed, $\boldsymbol{H}$ plays the role of a "post-processing" operator providing the model equivalents of the observed parameters. Since the ECMWF forecast model is spectral and the control variable is the vector of spectral model variables, the operator $\boldsymbol{H}$ is the product of several distinct operators. These include the inverse spectral transform ($\boldsymbol{H}_{it}$) which provides the model values on the gaussian grid at all model levels and a horizontal interpolation ($\boldsymbol{H}_{hi}$) from grid points to observation points for all the model levels. The current version of the ECMWF 3D variational analysis provides two interpolation options: (1) bi-linear (4 points), and (2) bi-cubic (12 points). The horizontal interpolation is common for all the observation types. The next step is to perform the vertical interpolation ($\boldsymbol{H}_{vi}$) to get the model equivalents of the observed quantities at the observed levels. $\boldsymbol{H}_{vi}$, unlike $\boldsymbol{H}_{hr}$, will differ for different parameters.

Given the above decomposition, the operator $\boldsymbol{H}$ can be written as product of the individual operators:

$$\boldsymbol{Hx} = \boldsymbol{H}_{vi}\boldsymbol{H}_{hi}\boldsymbol{H}_{it}(x)$$

The evaluation of the cost function is straightforward, given the observed departures and the inverse of the matrix $\boldsymbol{O}$.

VERTICAL UPPER AIR OBSERVATION OPERATORS

The following vertical interpolation techniques are employed:

- wind: linear in ln(p) on full model levels but quadratic at the two top levels
- geopotential: as for wind but the interpolated quantity is the departure from ICAO standard atmosphere

- humidity: linear in p on full model levels
- temperature: linear in p on full model levels

The integration of the hydrostatic equation used for the geopotential operator depends on both temperature and specific humidity.

SURFACE OBSERVATION OPERATORS

The vertical gradient of the model variables varies strongly in the lowest part of PBL, where flow changes are induced on very short time and space scales, due to physical factors such as turbulence and terrain characteristics. In such a situation the operator H_{vi} should take account of the model surface layer. The operators for 10m wind and 2m temperature are based on Monin-Obukhov similarity theory, whereas in the case of 2m relative humidity it is a somewhat simpler (still highly non-linear) operator. The similarity functions match the physical parametrization of the forecast model.

The post-processed values of 10m wind (u^{pp}_{10m},v^{pp}_{10m}) and 2m temperature (t^{pp}_{2m}) are functions of temperature (T_l), specific humidity (Q_l) and wind components (U_l,V_l) at the lowest model level. The surface pressure (p_s) and temperature (T_s) are also included as well as the roughness length (Z_o), vegetation (C_v), soil wetness of the first model soil layer (W_s), water content of the skin reservoir (W_l) and snow (S_n), or formally written:

$$(u^{pp}_{10m},v^{pp}_{10m},T^{pp}_{2m}) = f(T_l,Q_l,U_l,V_l,p_s,T_s,Z_o,C_v,W_s,W_l,S_n)$$

On the assumption that the relative humidity is constant in the model's surface layer, the post-processed 2m relative humidity (RH^{pp}_{2m}) is assigned the value of the lowest model level which is a function of the lowest model level pressure (p_l), specific humidity (Q_l) and temperature (T_l):

$$RH^{pp}_{2m} = f(p_l,Q_l,T_l)$$

4.2 TOVS cleared radiances

The computation of the TOVS observation cost function is organized like that for conventional data (previous section), with the addition that, for TOVS, horizontal observation error correlations are taken into account (*Pailleux et al.*, 1991). However, data from different satellites are assumed to be uncorrelated. Data from different retrieval types are also assumed to be uncorrelated. Thus it has been possible to separate the observations in several de-correlated sets. With large numbers of TOVS it becomes necessary to split the sets further. Technical reasons limit the number of members in each set to a couple of hundred (currently we use 192). The radiance observation error correlation function is specified to be Gaussian with a length scale of 350 km. Half of the observation error only is assumed to be correlated.

The forward operator $\boldsymbol{H}$, is the product of all the operations necessary to go from the control variable $\boldsymbol{x}$ to model radiances at observation points. The operator $\boldsymbol{H}$ is continuous in $\boldsymbol{x}$. It may be linear or nonlinear and it ought to be differentiable in general but it does not have to be differentiable for *all* values of $\boldsymbol{x}$. Linear interpolation is for example differentiable between model levels but not differentiable exactly *at* a model level.

The chain of operators in the TOVS forward and adjoint calculations is shown schematically in Fig. 16. It starts with a change of variable from the control variable to model spectral variables (see section 3), followed by the inverse spectral transforms to obtain grid-point data of temperature and specific humidity on the model's Gaussian grid. The model grid-point data are then interpolated in the horizontal with a 12-point bi-cubic interpolator and in the vertical to 40 pressure levels, assuming linearity of $\boldsymbol{T}$ and $\boldsymbol{q}$ in $\boldsymbol{p}$. The radiative transfer model (*Eyre*, 1991) is formulated in terms of $\boldsymbol{T}$ and $\ln \boldsymbol{q}$ on 40 fixed pressure levels from 1000 to 0.1 hPa. The calculations are carried out from $p = p(z_m)$ to 0.1 hPa, where z_m is the elevation of the TOVS locations, as given in the TOVS reports. The radiative transfer model also uses T_s and T_2m. As the control variable is currently limited to model level quantities we must either extrapolate the available information to 0.1 hPa or bring in auxiliary information from an external source. We have chosen to bring in the 1D-Var retrieved temperature at the surface (T_s) and in the stratosphere above the top of the model (7.3 hPa). We use the temperature at the lowest model level in place of T_2m.

For humidity we discard the model variables above 300 hPa. They are replaced with a constant value of $\boldsymbol{q}$ above 70 hPa and extrapolated according to an empirical power-law between 300 and 70 hPa.

Once model radiances have been computed, the cost function and its gradient with respect to radiances can be calculated. Then the adjoint operators are applied in the reverse order (Fig.16) to yield the gradient of the cost-function with respect to the control variable.

The forward calculation for J_o was validated by carrying out radiance calculations at all TOVS locations for a given six-hour forecast. RMS and bias of the observed departures from the calculated radiances were compared with the result of operational radiance computations (with the same radiative transfer model).

The adjoint calculations of Fig. 16 were validated with the so called gradient test, which tests that the adjoint is fully consistent with the forward operators.

Consider the univariate temperature analysis of *one* radiance datum at *one* location. Assume that temperature forecast errors are constant in the vertical. With the linear approximation the temperature analysis increment $(T_a - T_b)$, is given by:

$$T_a - T_b = BR'^T \, (R'BR'^T + O)^{-1} \, [y - R(x)] \tag{54}$$

where:

R' is $\partial R/\partial T_i$, and T_i is temperature at model level i

B is the temperature covariance matrix

O is the observation error variance of the channel in question

$R'BR'^T$ is the radiance equivalent of the forecast error variance

For the vertical distribution of the analysis increment (ignoring the amplitude), Eq. (54) simplifies to:

$$T_a - T_b \propto BR'^T \tag{55}$$

The matrix R' represents the sensitivity of a given TOVS channel to the temperature at discrete model levels. The matrix B describes how the observation increment is spread in the vertical. In the first validation experiment we simplify further and assume a diagonal B. This de-couples the analysis in the vertical, and we obtain a two-dimensional analysis. Hence, the analysis increment at each level should be proportional to R'.

Carrying out a full T63 3D-Var analysis, with the above simplifications, using the channel MSU-2 only, gives the vertical profile of analysis increment given in Fig. 17b. Comparing with R' for MSU-2 as shown in Fig. 17a we see that the analysis indeed gives the expected solution.

With a non-diagonal vertical correlation matrix, the theoretical analysis increment at a given level depends on R' at all other levels, according to Eq. (55). The theoretical analysis increment with the full B matrix is shown in Fig. 18a (thick line - the thin line represents 1D-Var and is here as a reference to be compared with the actual analysis increment in Fig. 18b). The two are virtually identical, which fully validates the vertical aspects of 3D-Var.

4.3 Scatterometer data

The direct use of scatterometer data in 4D-Var is described in *Thépaut et al.* (1993a). For this purpose the model function developed for ESA has been used (*Stoffelen*, 1992). An open question is whether to use directly the σ_o or to use retrieved ambiguous wind. A particularity of these data is the low level of instrumental noise: the data lies very close to a surface in the 3D σ_o space which is a simple (but highly non linearised 2-folded) mapping from wind space. The errors in the mapping are dominant and are easier to specify in wind space than in σ_o space. In particular in σ_o space we would have to parametrize errors which are very close to the cone and consistent with the statistics derived from collocations in wind space.

We are therefore investigating the specification of cost function in wind space which has 2 minima corresponding to the 2 ambiguities and which is consistent with the statistics.

5. 3D-VAR RESULTS

5.1 Single observation experiments

The multivariate balance imposed by J_b as described in section 3.3 provides a mass wind coupling over the whole globe. In the examples which follow all vertical modes are used in the separation between Rossby and gravity waves. Horizontal modes used are those corresponding to vertical modes 1 through 7. Beyond vertical mode 7 the same set of horizontal modes (those associated with vertical mode 7) are used for all higher modes. Note that there is no univariate component in this example. Fig. 19 shows the response to an isolated observation at 60° N of a) height, b) zonal wind, and c) meridional wind. In each case the analysis increment is near geostrophic. The horizontal scale is determined by the horizontal structure functions and the vertical spread by the vertical structure functions as described before.

One of the limitations of the ECMWF OI analysis is the lack of mass/wind balance as one approaches the equator - the scheme becomes univariate at the equator. Fig. 20 shows the response to an isolated observation at the equator of a) positive zonal wind, b) negative zonal wind, and c) a southerly meridional wind. The variational analysis has a strong mass-wind balance even at 0°. Note, however, the absence of a Kelvin wave response, in the current formulation these are taken as "fast" modes and assigned relatively large errors. As *Parrish* (1988) and *Daley* (1993) point out, Rossby modes imply a negative u, ϕ correlation at the equator, whereas Kelvin modes imply a positive u, ϕ correlation at the equator. The addition of Kelvin modes in the "slow" term of J_b will considerably reduce the u, ϕ correlations at the equator (*Parrish*, 1988).

Fig. 21 shows the linear balance response to an isolated zonal wind observation at 0°N. Compare this with the Hough balance shown in Fig. 20a. The similarity with the linear balance solution is, of course, affected by the number of vertical modes used in the Rossby/gravity separation - in this case 7. Closest similarity with the linear balance solution is obtained when the external mode is used for the separation as Hough balance becomes identical to the linear balance in the limit of infinite equivalent depth. One can expect there to be sensitivity of the increments to the details of the Rossby/gravity/univariate separation, again this will have to be an area of further research.

Finally, it is worth noting the effect of varying the ε parameter introduced in section 3.4. In all the above $\varepsilon = 0.1$, which implies that 10% of the error variance lies in the gravity wave part of the fields. Fig. 22a shows the effect, for a single zonal wind observation at 0°N of $\varepsilon = 0.5$, and Fig. 22b of $\varepsilon = 0.9$, c.f. Fig. 20a which is the $\varepsilon = 0.1$ case. The balance changes from near geostrophic with $\varepsilon = 0.1$, to almost entirely ageostrophic with $\varepsilon = 0.9$. As discussed in section 3.4, the ε parameter is analogous to the OI formulation in which one assumes that a certain percentage of the variance is described by the divergent component of the wind. *Daley's* (1983) experiments indicated that 10% was a reasonable figure for this, and following further experimentation by *Undén* (1989) this is the value used operationally by the ECMWF OI. Its role is slightly different with the variational analysis and this is another area which may benefit from a closer study.

5.2 Assimilation experiments

Several 2 week assimilations have been performed to evaluate the strength and weakness of the system. The use of satellite data is documented in *Andersson et al.* (1993). The 3D-Var T63L19 experiment 'caa' is discussed which uses non separable structure functions as described in section 3. It was run for a two week period, together with a parallel control, 'bxw', (1D-Var T63L19). The performance of both analyses is very similar, as is demonstrated by the forecast scores. A synoptic study will be discussed at day ten of the assimilation and some 7 day mean results are shown.

SURFACE RESULTS 00Z 21 NOVEMBER 1992

This date was chosen because there was a deep low in the Northern Atlantic and it was well into the data assimilation period. Fig. 23(a) shows the 1000 hPa for both experiments contoured on top of METEOSAT 11 micron imagery for the European and Atlantic region. The general impression is there are very minor differences between both experiments. Fig. 23(b) zooms in on the Atlantic low and there is very close agreement in both analysis systems.

Turning to the Southern Hemisphere one might think that there may be larger differences between the experiments. Fig. 23(c) shows the South Atlantic with the coast of Antarctica at the bottom for the figure. The differences are all little larger but it is difficult to decide using the cloud imagery which is the more correct system because there are not major phase differences in any synoptic system.

MEAN ANALYSES

After five days of data assimilation it is reasonable to assume that both systems would be independent of the initial state. Hence the next seven days at 00Z were averaged for all model variables on the 19 model levels on a 5 by 2.5 degree grid and the mean analysis were computed.

ANALYSES ON MODEL LEVELS

Figure 24(a) shows the mean temperature for both experiments on model level 11 which is close to 500 hPa. Both experiments are very similar but the 3D-Var has a little more amplitude in both troughs and ridges.

Upper level temperatures (near 50 hPa, model level 2) are shown in Fig. 24(b). There are now larger differences between the experiment than lower down but the results in the northern hemisphere are still close. It is in the Southern Hemisphere that differences of up to five degrees can be observed.

The lowest model level humidities are shown in figure 24(c). AT these levels there does not appear to be any systematic differences in humidity.

MEAN CROSS SECTIONS

The visualization of analysis differences can be seen using zonal cross sections. The following sections are all taken between 60 degrees south and 60 degrees north at zero longitude from the surface to 10 hPa.

COMPARISON BETWEEN 3D-VAR AND OI T63/L19 CONTROL

Fig. 25(a) shows a mean temperature section for both experiments. There is good agreement with the maximum differences less than two degrees except in the stratosphere.

The mean wind speed section is shown in Fig. 25(b). The north and south jets are in good agreement but there is a small difference in the sub-tropical jet near profile 15. The 3D-Var analysis reduces the area covered by the jet core.

Humidity sections are displayed in Fig. 25(c) up to 500 hPa. There is good agreement in regions outside the Southern Hemispheric sub tropics. In this region where there is a descending branch of the Hadley circulation the 3D-Var appears to be a little better than the OI control. The drier air descends to a lower level.

COMPARISON BETWEEN 3D-VAR AND OI T213/L31 OPERATIONS

Figs. 26(a), (b) and (c) are very similar to Figs. 25(a), (b) and (c). The major difference is a small increase in the jet speed as expected with the increased resolution in the T213/L31 operational model.

OBJECTIVE SCORES

Fig. 27 presents the scatter diagram of the RMS of the 500 hPa forecast error issued from 3D-Var (horizontal) and from OI-1D-Var (vertical) verified against the operational analyses (panel a: Northern Hemisphere, panel b: Europe, panel c: Southern Hemisphere).

The Northern Hemisphere scores are neutral in the early medium range and slightly negative at day 6. This comes mainly from a single case which, while looking at the synoptics (Fig. 28), is very comparable to the OI forecasts. Looking at the difference maps with the verifying analysis (Fig. 29), they are remarkably similar over Europe and in the Pacific, while slightly larger for 3D-Var. However, they are larger over North Canada where the jet comes south to the Hudson Bay in the 3D-Var forecast, while staying North in both ops and the OI forecast.

The European scores are better at all ranges as are the Southern Hemisphere scores. At the 1000 hPa (not shown) they display essentially the same story.

INCREMENTAL 3D-VAR

One week assimilation has been performed at truncation T106 with increments at T63 on the same period as above. The results are very comparable to the previous ones, thus confirming in practice the validity of the incremental approach.

6. 4D-VAR EXPERIMENTS

Several 4D-Var experiments have been described in *Andersson et al.* (1993) using TOVS data, in *Thépaut et al.* (1993b) using conventional data and in *Thépaut et al.* (1993a) using scatterometer data. In summary, they showed the ability of 4D-Var to generate flow dependent and baroclinic structure functions and to extract wind information from the TOVS humidity channel. These experiments have been performed using the original 4D-Var formulation; here we shall describe some results obtained using the incremental approach.

A first 4D-VAR experiment has been performed at T106 resolution and with full physics (increments at T63 with an adiabatic model). Three 6 hour cycles have been carried out from 15/10/92 00 UTC to 15/10/92 12 UTC. Only conventional observations (including SATOB winds) were used. A control 3D-VAR experiment has also been run under the same conditions. J_b is formulated with separable structure functions in both cases. Two ten day forecasts have been performed from the two assimilations and compared with the operational forecast (OPS). In Fig. 30 the 500 hPa anomaly correlation scores are presented, of the forecasts performed from 4D-VAR (left panel) and 3D-VAR (right panel) for the Northern (top) and the Southern hemisphere (bottom). If both forecasts behave similarly in the Southern hemisphere, one can notice a clear advantage in favour of 4D-VAR in the Northern Hemisphere (almost a day at the 60% level).

If we now look at the 500 hPa fields averaged over of the last 5 days of the forecasts (Fig. 31), we can clearly see that 4D-VAR (top left) looks more similar to the analysis (bottom left) than 3D-VAR (bottom right) or OPS (top right) specially in the Atlantic region where both 3D-VAR and OPS generate a cut-off 20W of Spain. The three forecasts have the same signature in the Pacific area.

This (single) result confirms the 4D-Var potential and validates the incremental approach. Using separable structure functions, a parallel five day assimilation was then run (6 hour 4D-VAR and OI) at a resolution of T106 from 10/03/93 12UTC to 14/03/93 12 UTC, again using conventional observations only. The outcome of the experiments (these last five cases and the first one described above) is the following:

- In the Northern hemisphere, we have 2 neutral cases, two slightly positive cases (in favour of 4D-VAR), one clearly negative case and one clearly positive.

- In the Southern hemisphere, we have two neutral cases, 1 slightly positive case, 1 slightly negative case and two clearly positive cases.

Fig. 32 represents the 500 hPa scores averaged over the 6 cases, for the Northern and Southern hemispheres.

The main conclusion from these early experiments is that the incremental 4D-VAR approach works. From a pure theoretical point of view it is not surprising since the formulation is only a simplification of the full assimilation problem, but these six cases validate the robustness of the method. In the Southern hemisphere, benefit is taken from the dynamical consistency of the analysis in data-void areas (even if the temporal window is only 6 hours on those cases).

During these experiments, several weaknesses (and bugs!) in the practical implementation of the incremental 4D-VAR have been identified.

- At the time of the experiments and for practical reasons, the models used to perform the assimilation (IFS) and the 6 hour forecast (SPM) were not identical. In particular, the assimilating model was still under validation with known small problems in the physics (e.g. no shallow convection).

- The control of the gravity waves in the analysed increments was not handled optimally.

- We have also spotted a problem of convergence in the scheme. The choice for the experiments was to perform 15 iterations for each inner minimization, and two external updates of the trajectory. We have some evidence that updating more often the trajectory will lead to more robustness in the definition of ***R*** but the actual number of updates remains to be tuned, bearing in mind some trade-off with computer resources. Moreover, we have realised that increasing to 25 the number of iterations could decrease the final cost function by an additional 30%.

Obviously some tuning/refining exercise is needed to improve on the method. In particular, the inclusion of satellite observations, the improvement of the assimilating model and the study of the optimal temporal assimilation window (6 h, 12 h or 24 h) should give us some insight about a possible pre-operational scenario.

7. CONCLUSIONS AND PERSPECTIVE

In section 2 we have presented the incremental approach for 3D and 4D-Var and we have related the approximations so performed to simplification of the Ricatti equation (transport in time of the covariances) of the Kalman-Bucy filter. This formulation has the advantage of allowing trade-off between the 4D-Var CPU and memory costs (only memory for 3D-Var) and the scientific improvements. If the scientific outcome of sections 5 and 6 is confirmed, operational implementation of 3D-Var should be feasible in early 1994 and of 4D-Var in the foreseeable future.

Here we have not presented some work which took place on preconditioning (*Courtier et al.*, 1993). In order to implement some of the conditioning improvements, interactions have been necessary with INRIA in order to make M1QN3 more flexible, the minimization algorithm used in all experiments reported here (*Gilbert and Lemaréchal*, 1989). These results combined with ideas provided by Nocedal (personal communication) should allow us to compute an approximation of the Hessian and of its inverse, the covariance matrix of analysis error. Scientific evaluation is necessary to evaluate the potential of the approach. Following the same ideas, it should be possible to implement a simplified Kalman filter which uses the singular vector concept at the basis of the computation of the initial perturbations of the ensemble prediction system. Further theoretical work is necessary but the ideas seem promising.

In sections 3 and 4 we have shown that some separability in the structure functions are easy to implement in 3D-Var, which was a long outstanding issue in the OI framework (solved recently by *Bartello and Mitchell*, 1992). This is crucial if one attempts to extract optimally and simultaneously temperature and wind information from the observations. In the specification of the statistics, we have concentrated up to now on the balanced part and work has just started for the unbalanced part. We are, however, confident that the current formulation of the background term is robust enough for most of the foreseen enhancements. Currently the treatment of the balanced part assumes that the Rossby modes are determined by vorticity. If this is true for small scales, this is not the case for the planetary scales. The natural evolution is to treat those scales (or even those explicit in the mixed implicit/explicit initialisation scheme) directly in Hough space. However, we do not consider this as crucial for operational implementation.

The main conclusion of section 5 is that 3D-Var in its current formulation has a quality similar to OI. This is very promising since we are really in the early stages of the tuning process. From section 6 we have seen that the incremental 4D-Var is very encouraging in terms of the quality of the medium-range forecasts.

References

Andersson, E., J. Pailleux, J-N. Thépaut, J.R. Eyre, A.P. McNally, G.A. Kelly and P. Courtier, 1993: Use of cloud-cleared radiance in three/four-dimensional variational data assimilation. Submitted to Q.J.R.Meteorol.Soc.

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

Charney, J.G., 1971: Geostrophic turbulence. Jour.Atm.Sci., 28, 1087-1095.

Courtier, P. and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. Tellus, 42A, 531-549.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1993: A strategy for operational implementation of 4D-VAR, using an incremental approach. Submitted to Q.J.R.Meteorol.Soc.

Daley, R., 1983: Spectral characteristics of the ECMWF objective analysis system. ECMWF Technical Report No. 40, 117 pp, ECMWF, Reading.

Daley, R., 1993: Atmospheric data assimilation on the equatorial beta plane. (Submitted for publication).

Eyre, J.R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. 1: Theory and simulation for TOVS. Q.J.R.Meteorol.Soc., 115, 1001-1026.

Eyre, J.R., 1991: A fast radiative transfer model for satellite sounding systems. ECMWF Technical Memorandum 176.

Eyre, J.R., 1992: A bias correction scheme for simulated TOVS brightness temperatures. ECMWF Technical Memorandum No. 186.

Gilbert, J.-Cl. and C. Lemaréchal, 1989: Some numerical experiments with variable storage quasi-Newton algorithms. Mathematical programming, B25, 407-435.

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.

Jazwinski, A.H., 1970: Stochastic processes and filtering theory. Academic Press, New York, 376 pp.

Li, Y., J.M. Navon, P. Courtier and P. Gauthier, 1993: Variational data assimilation with a semi-Lagrangian semi implicit global shallow-water equation model and its adjoint. Mon.Wea.Rev., 121, 1759-1769.

Lönnberg, P., 1988: Developments in the ECMWF analysis system. 1988 ECMWF seminar on data assimilation and the use of satellite data, pp 75-119.

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

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

Mitchell, H.L., 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.

Pailleux, J., W. Heckley, D. Vasiljevic, J-N. Thépaut, F. Rabier, C. Cardinali and E. Andersson, 1991: Development of a variational assimilation system. ECMWF Technical Memorandum 179, 51 pp.

Parrish, D.F., 1988: The introduction of Hough functions into optimal interpolation. Proceedings of the eighth conference on Numerical Weather Prediction, Baltimore, Md., Feb. 22-26. Amer.Meteor.Soc., Boston, MA, 191-196.

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. Mon.Wea.Rev., 120, 1747-1763.

Rabier, F., 1992: Assimilation variationnelle de données météorologiques en présence d'instabilité barocline. PHD dissertation Université Paris 6, available from the author.

Rabier, F., P. Courtier, J. Pailleux, O. Talagrand, D. Vasiljevic, 1993: A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis. Q.J.R.Meteorol.Soc., 119, 845-880.

Rostaing, N., S. Dalmas and A. Galligo, 1993: Automatic differentiation in Odyssée. To appear in Tellus.

Shaw, D., P. Lönnberg, A. Hollingsworth and P. Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF mass and wind analysis. Q.J.R.Meteorol.Soc., 113, 533-566.

Stoffelen, A.C.M. and D.L.T. Anderson, 1992: ERS-1 scatterometer calibration and validation activities at ECMWF: A. The quality and characteristics of the radar backscatter measurements. Proc. European 'International Space Year' Conference, Munich, Germany, 30 March - 4 April 1992.

Thépaut, J.-N. and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. Q.J.R. Meteorol.Soc., 117, 1225-1254. Also available from ECMWF as Technical Memorandum 178.

Thépaut, J.-N., R. Hoffman and P. Courtier, 1993a: Interactions of dynamics and observations in a four-dimensional variational assimilation. To appear in Mon.Wea.Rev.

Thépaut, J.-N., D. Vasiljevic, P. Courtier and J. Pailleux, 1993b: Variational assimilation of conventional observations with a multilevel primitive equation model. Q.J.Roy.Meteorol.Soc., 119, 153-186.

Undén, P., 1989: Tropical data assimilation and analysis of divergence. Mon.Wea.Rev., 117, 2485-2517.

Wergen, W., 1992: The effect of model errors in variational assimilation. Tellus, 44A, 297-313.

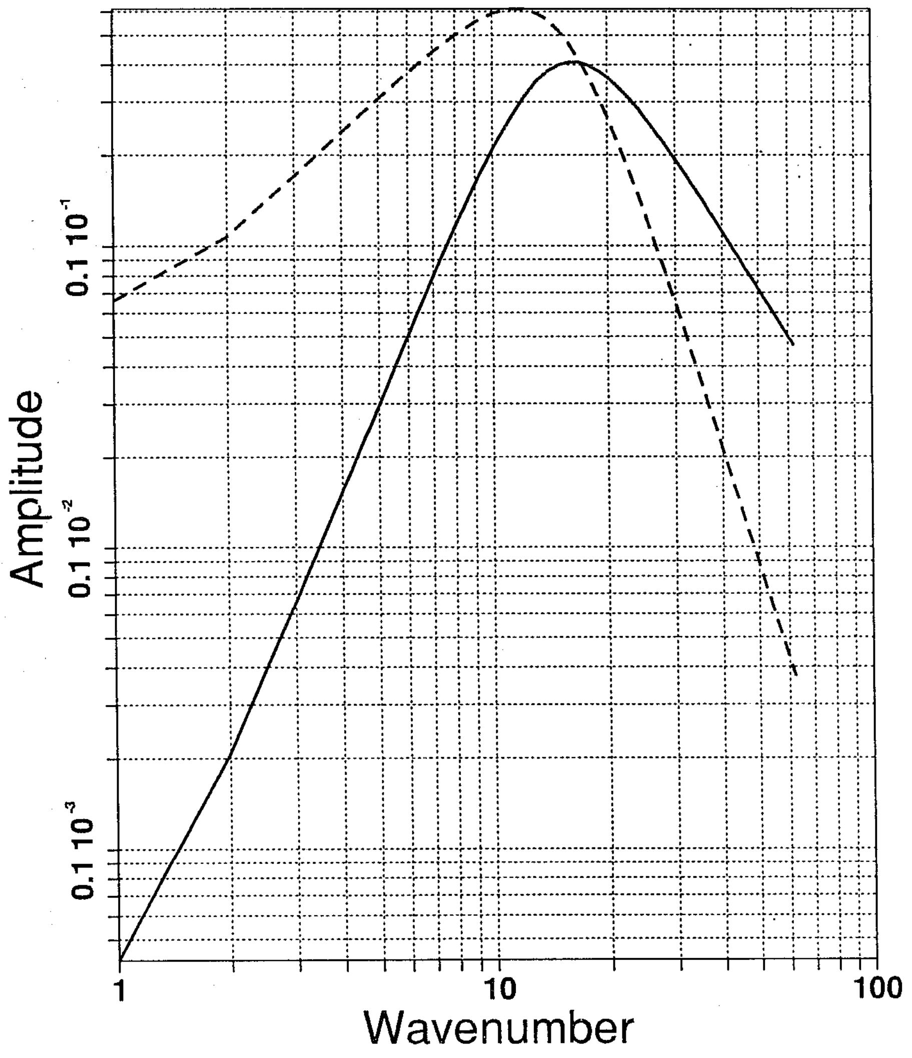

Fig. 1 Spectrum of the autocorrelation function of the wind errors (dashed) and height (solid). Parametric formulation.

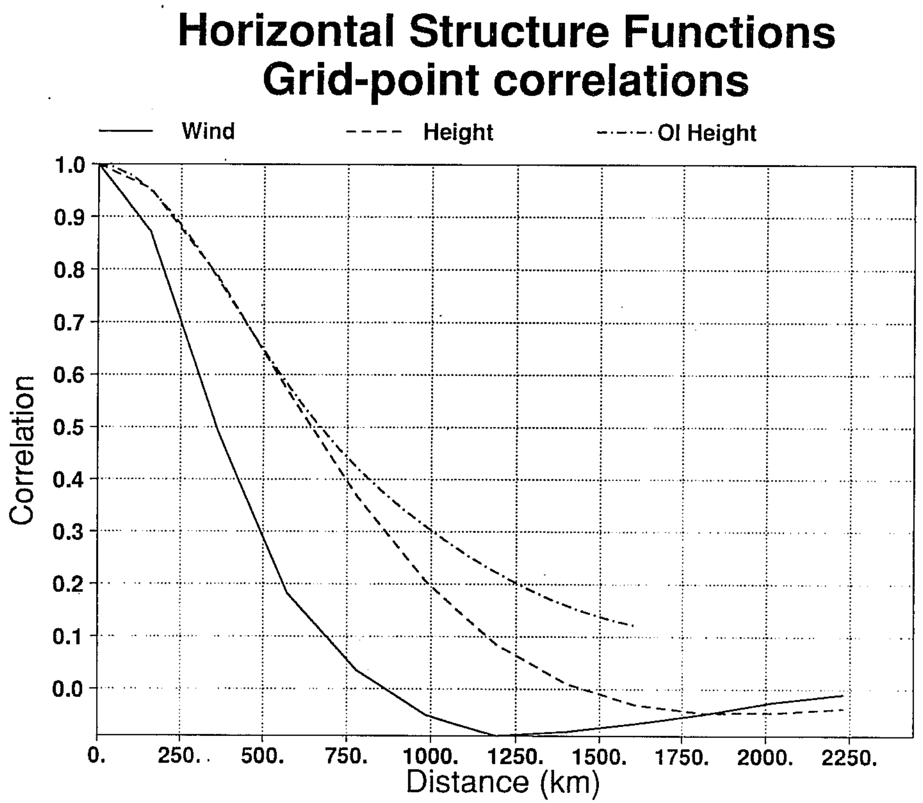

Fig. 2 Autocorrelation function of the geopotential errors. Solid and dashed line as specified in the parametric formulation of 3D-Var for wind and height, dot-dashed as generally used in OI.

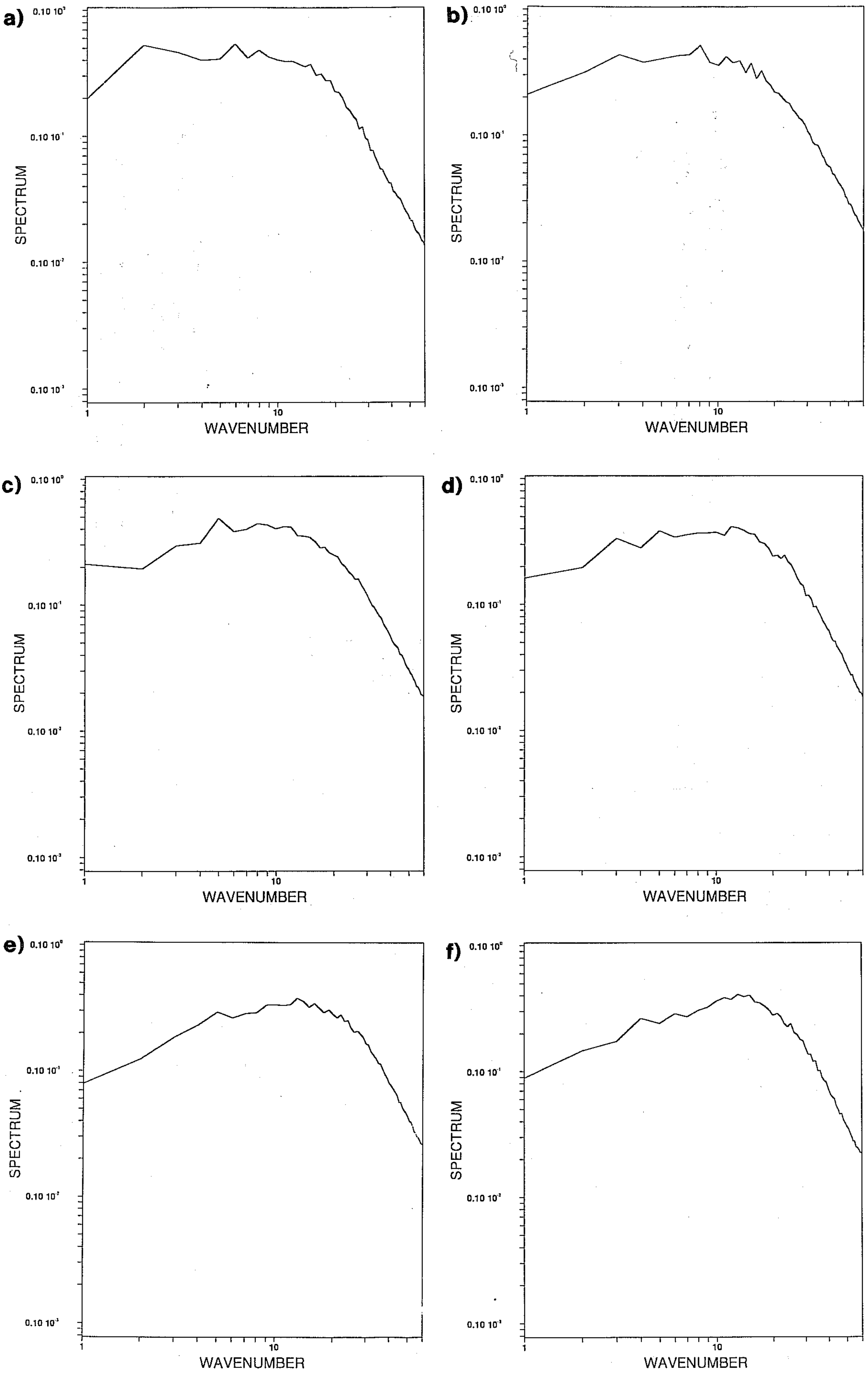

Fig. 3 Spectrum of the autocorrelation of forecast errors evaluated from differences with clear column radiance data a) MSU4 (70-100 hPa), b) HIRS4 (400 hPa), c) MSU2 (600-700 hPa), d) HIRS15 (700-800 hPa), e) HIRS13 (ground), f) HIRS11 (humidity channel, 700 hPa)

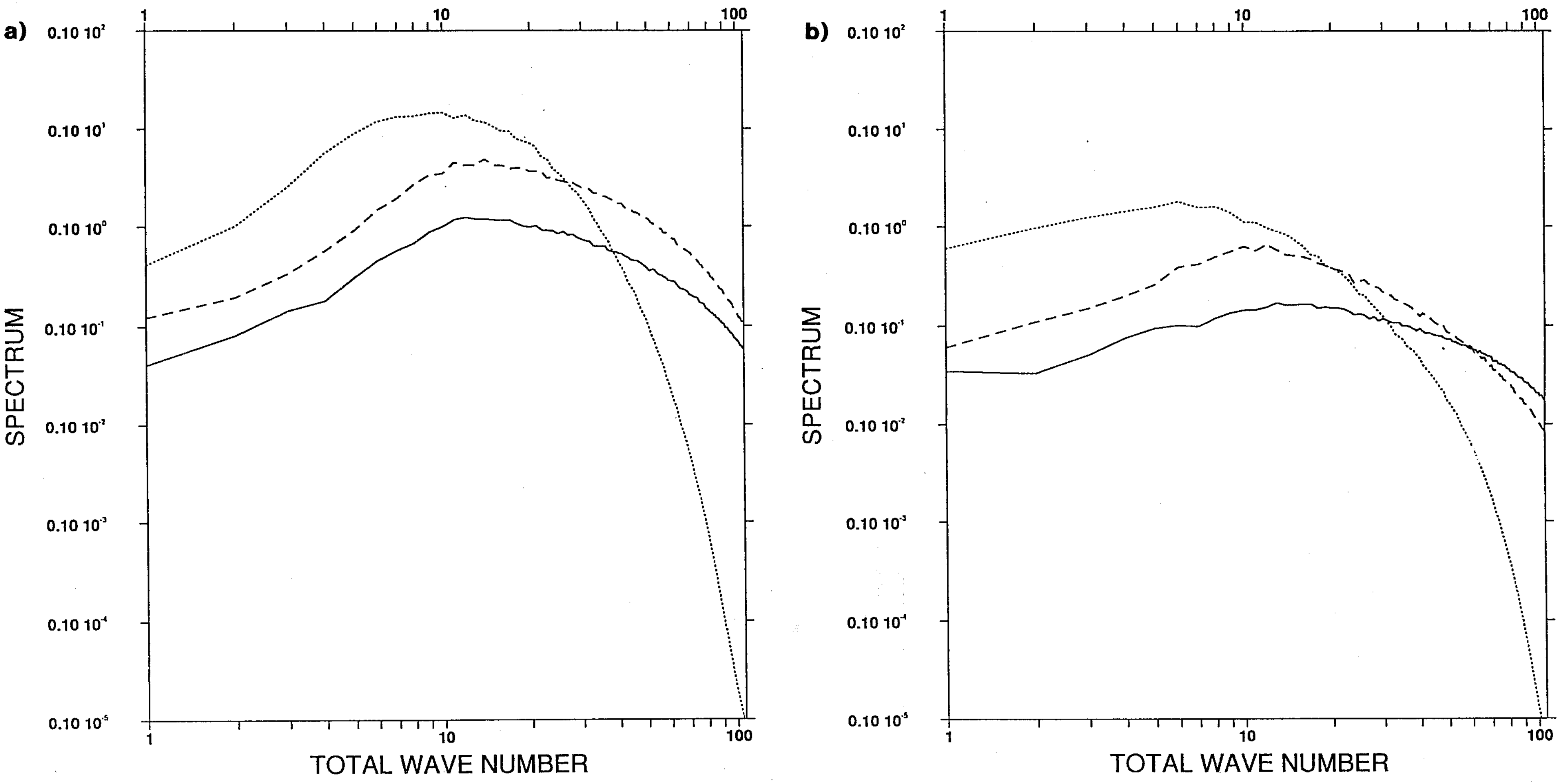

Fig. 4 Variance spectra of forecast error as evaluated using the NMC method of 24-48 hour forecast differences. Dotted line: model level 1 (10 hPa), dashed line: model level 18 (~ 500 hPa), solid line: model level 31 (surface). a) wind, b) temperature.

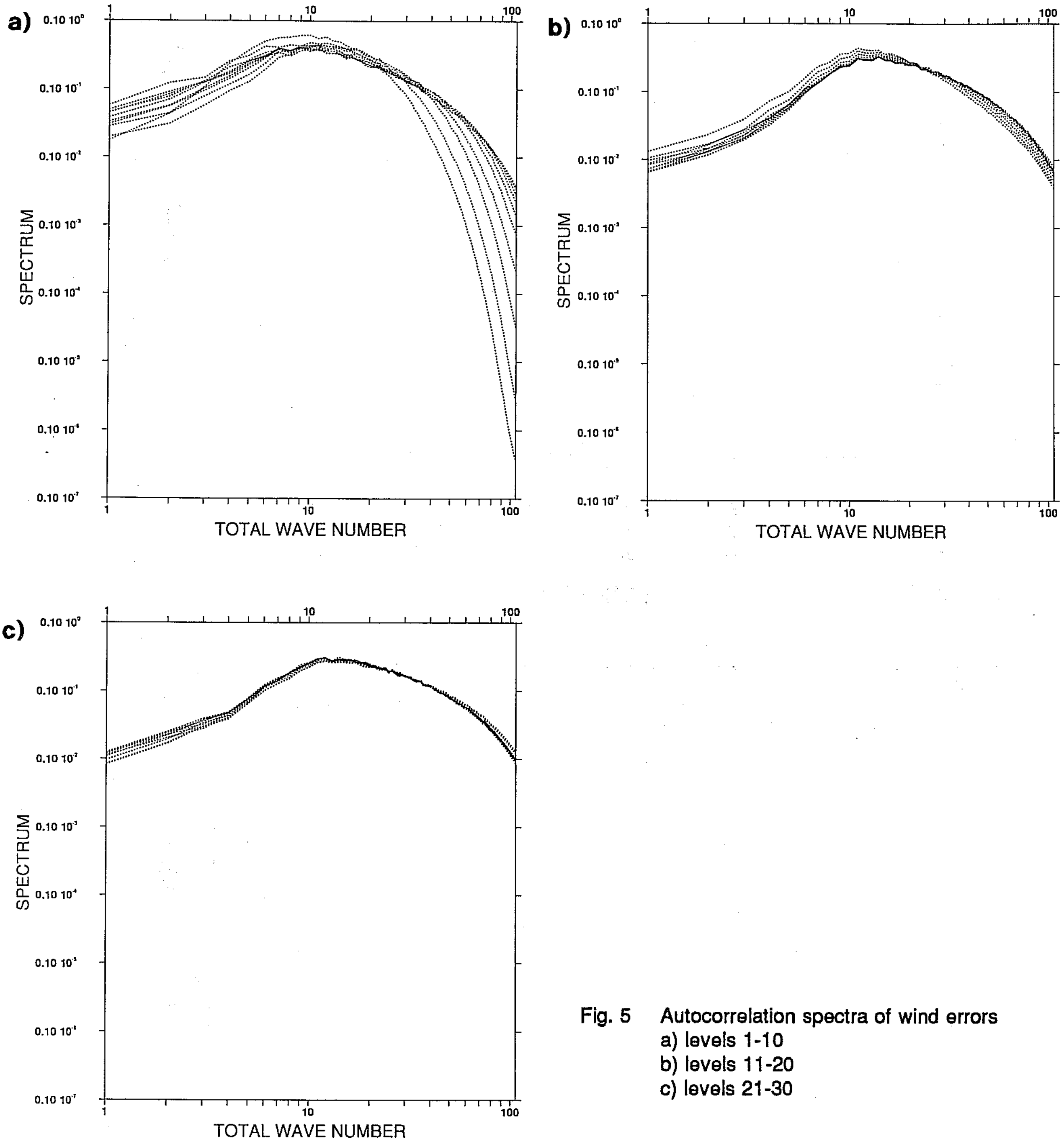

Fig. 5 Autocorrelation spectra of wind errors
a) levels 1-10
b) levels 11-20
c) levels 21-30

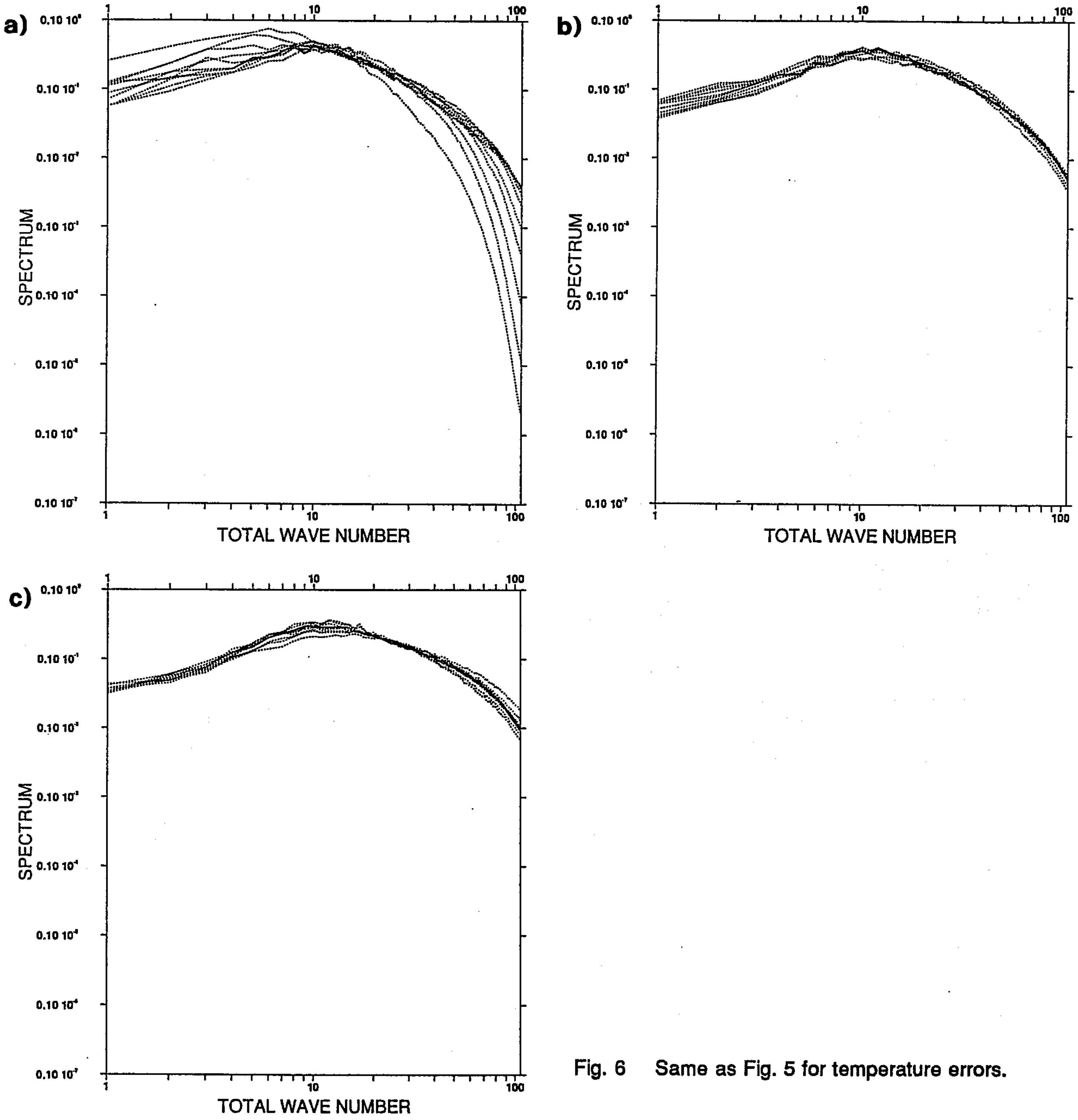

Fig. 6 Same as Fig. 5 for temperature errors.

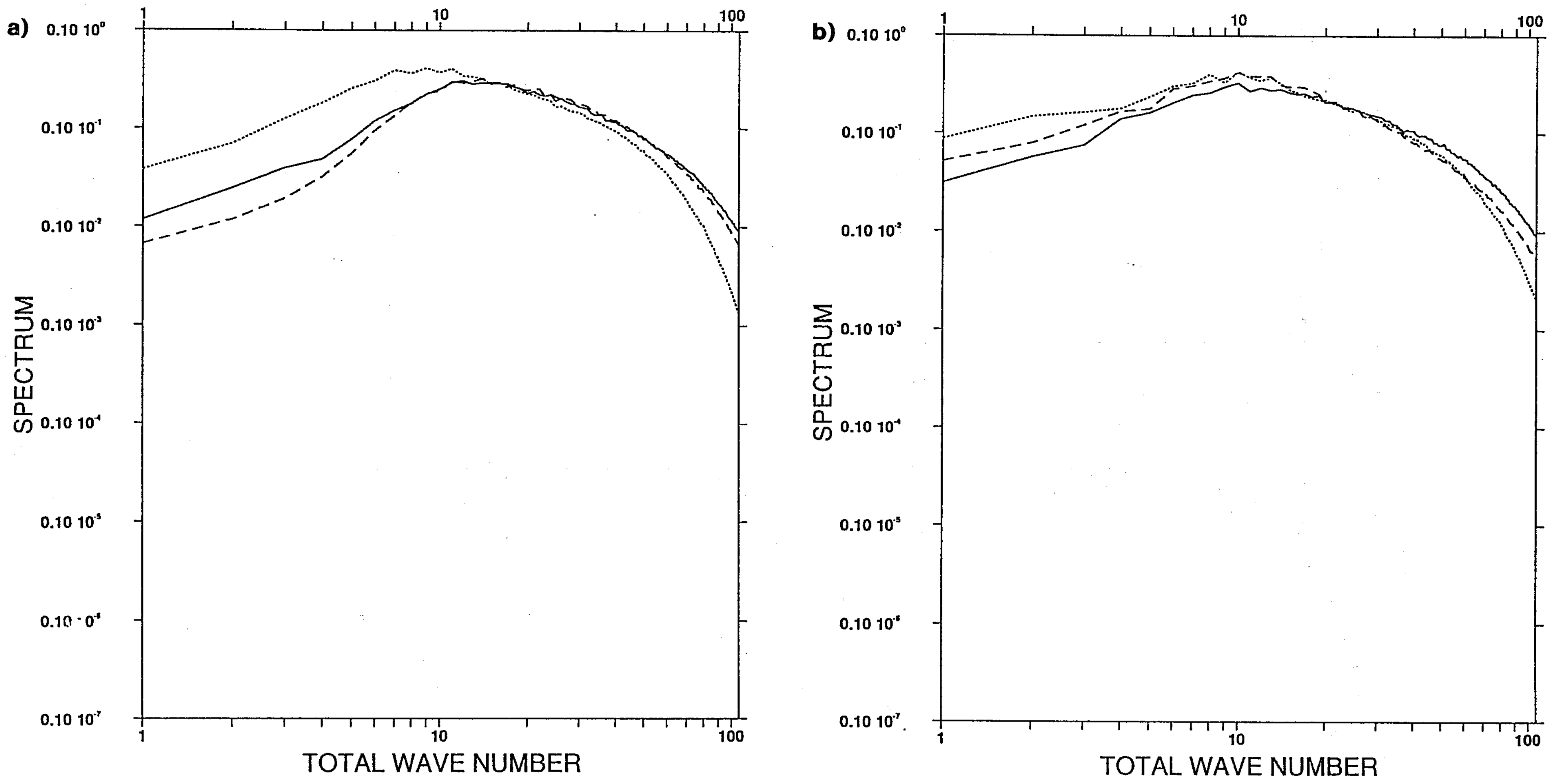

Fig. 7 Autocorrelation spectra of forecast errors. Dotted line: model level 6 (~ 100 hPa), dashed line: model level 16 (~ 400 hPa), solid line: model level 26 (~ 850 hPa). a) wind, b) temperature.

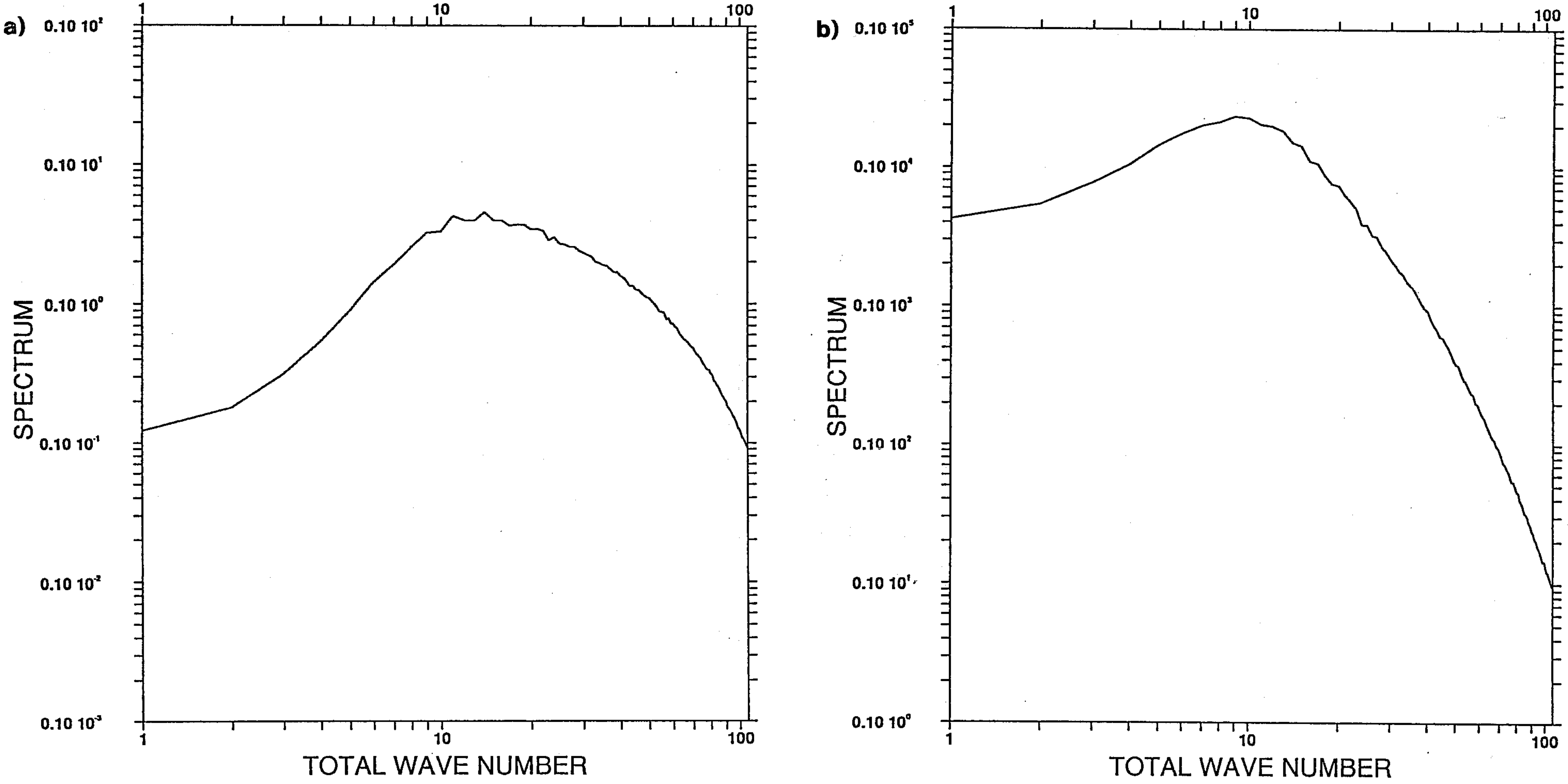

Fig. 8 Variance spectra of forecast errors at pressure level 500 hPa a) wind, b) geopotential

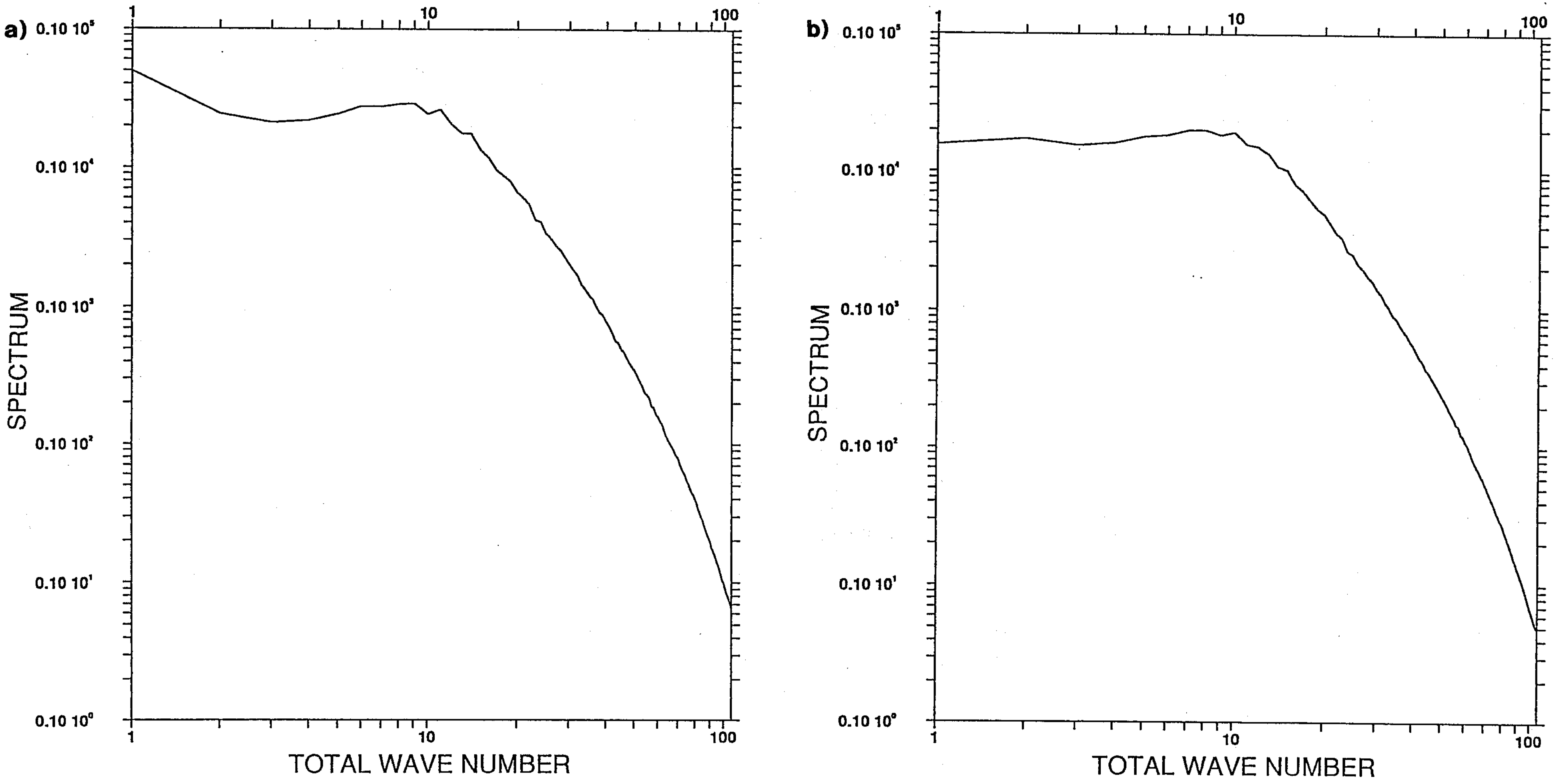

Fig. 9 Variance spectra of geopotential as obtained from wind spectrum a) using geostrophic balance on a f plane, b) using a linear balance relation

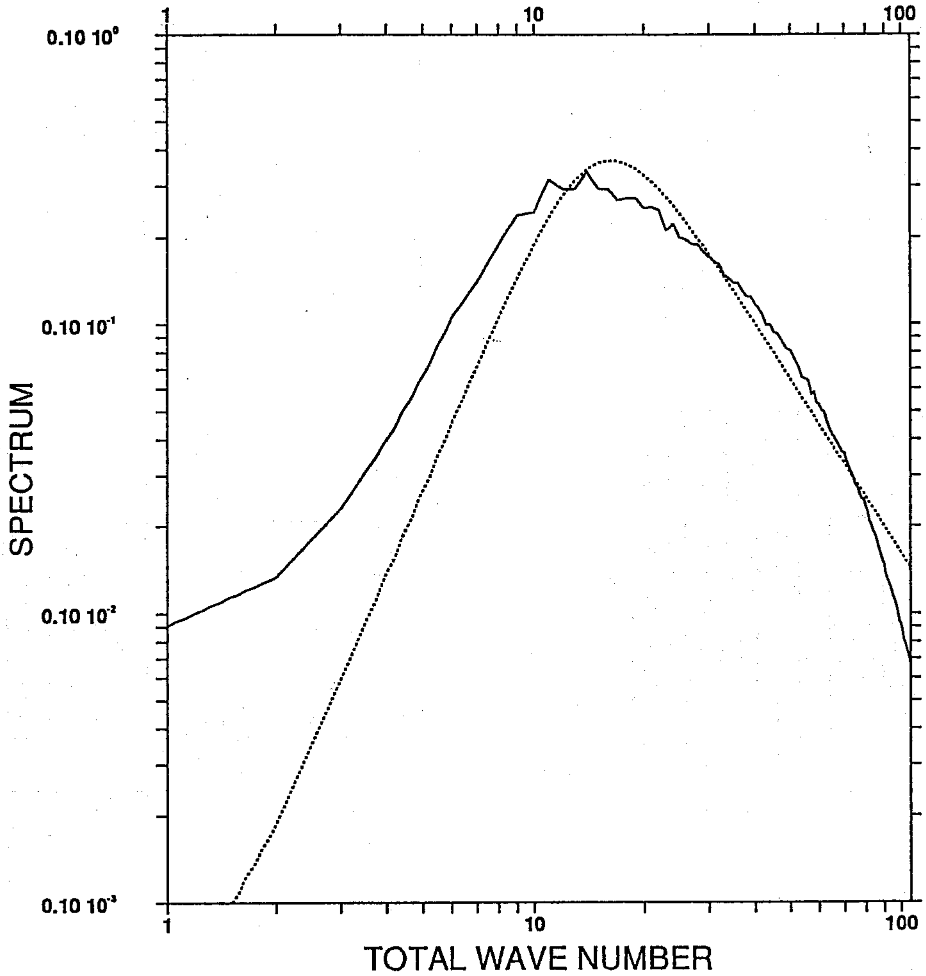

Fig. 10 Autocorrelation spectrum of the 500 hPa wind dotted: parametric formulation, solid: as obtained for NMC method.

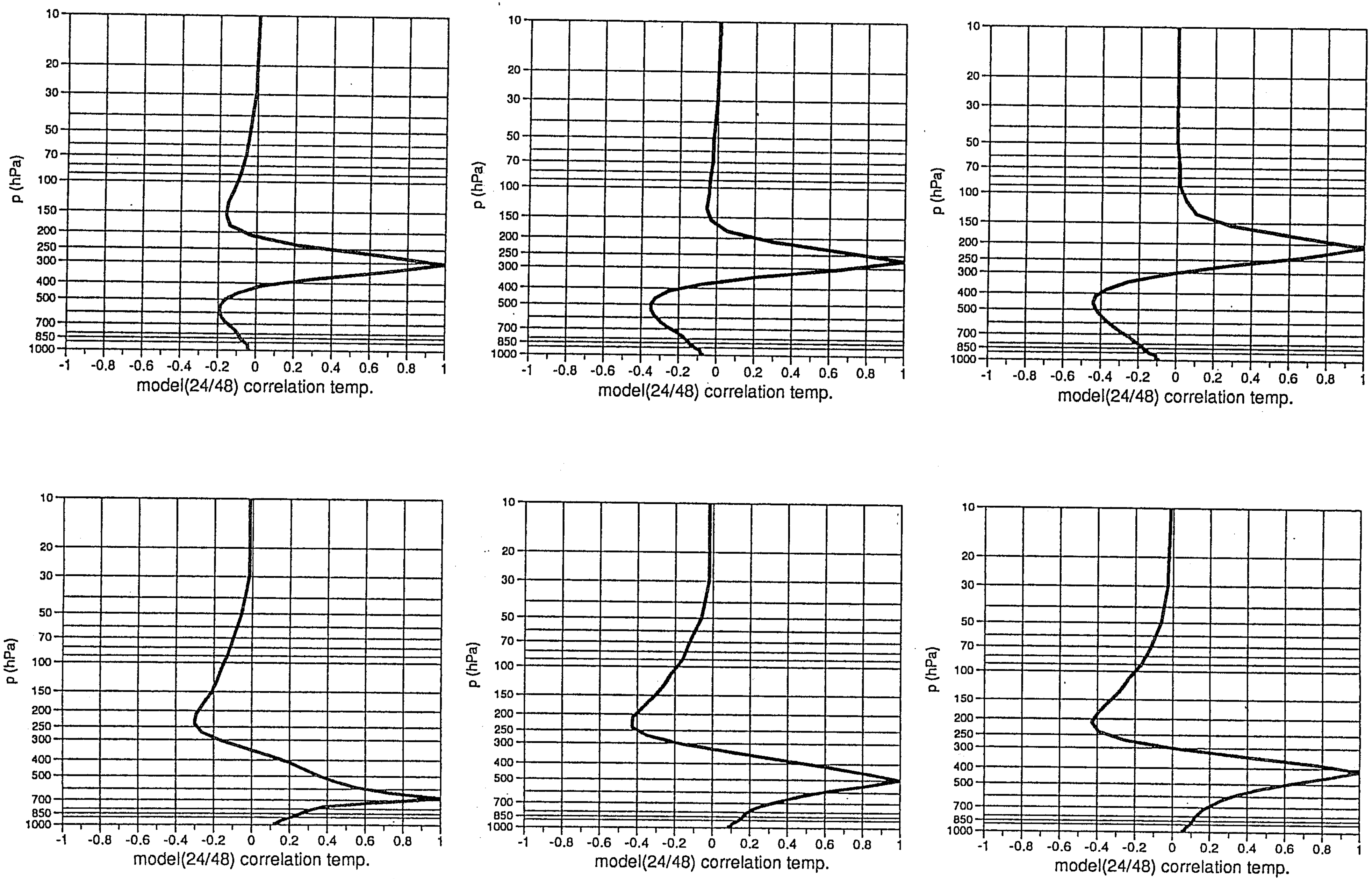

Fig. 11a) Vertical correlation of forecast errors as computed using NMC's approach: temperature.

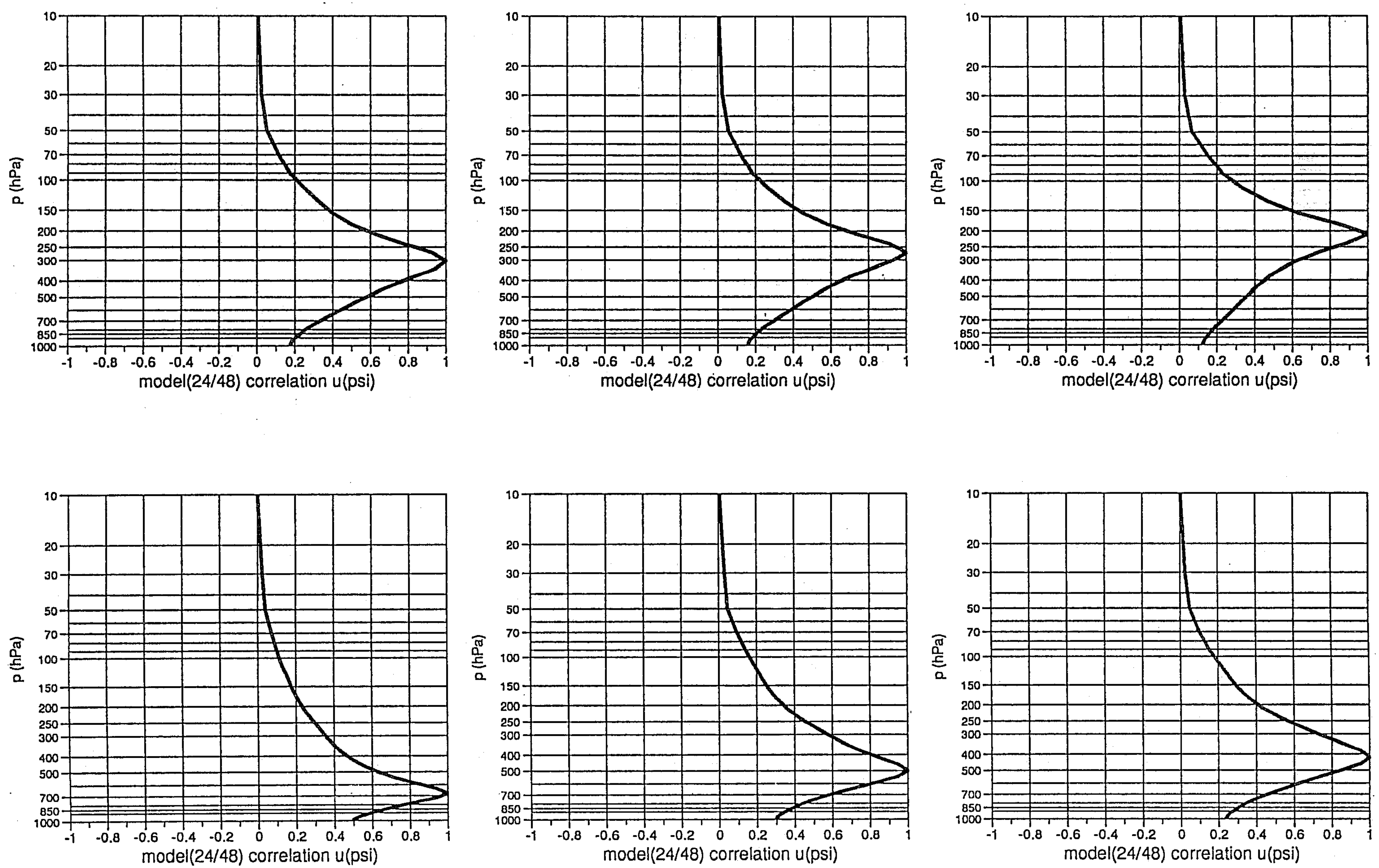

Fig. 11b) Vertical correlation of forecast errors as computed using NMC's approach: rotational part of the wind.

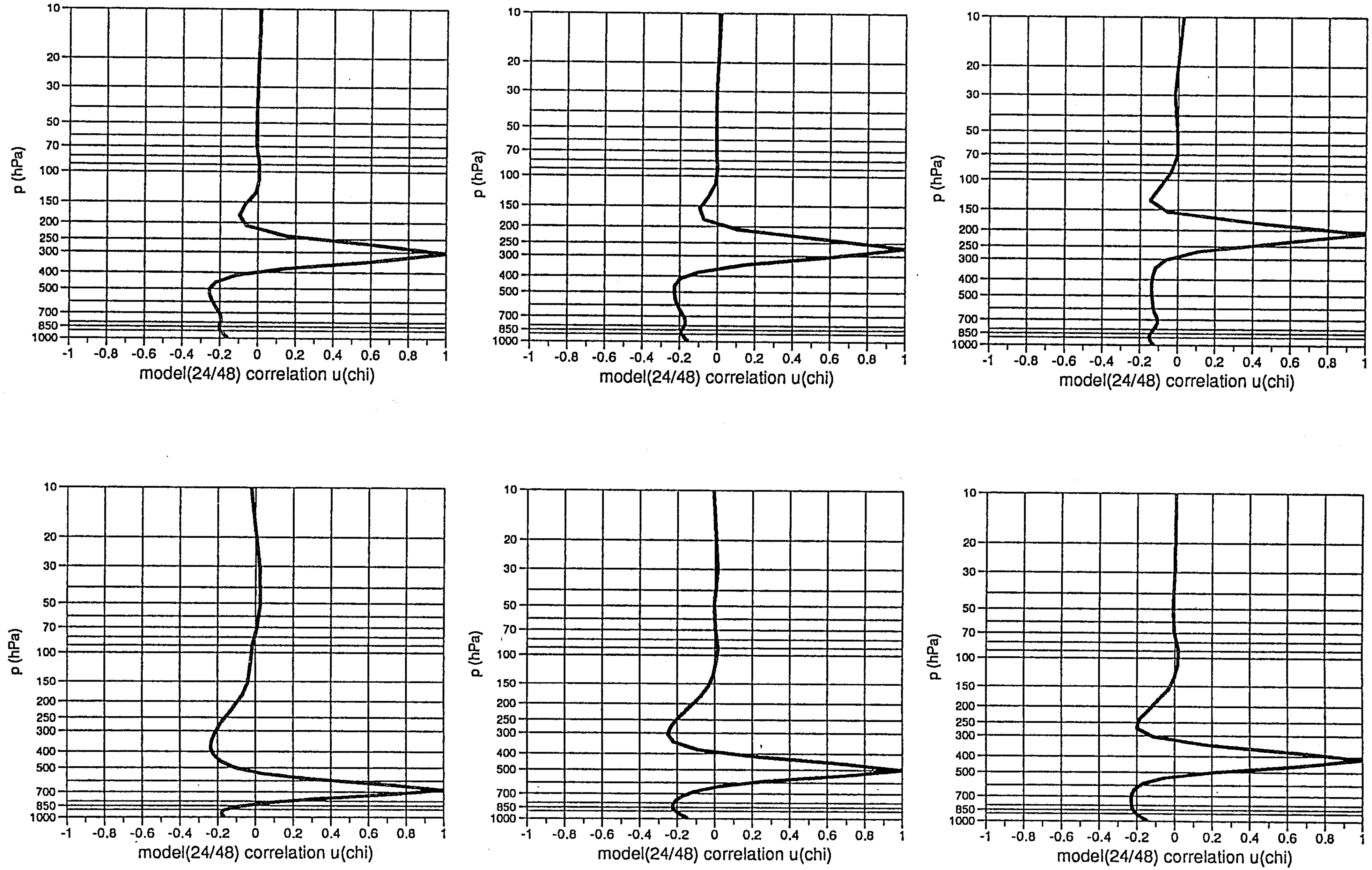

Fig. 11c) Vertical correlation of forecast errors as computed using NMC's approach: divergent part of the wind.

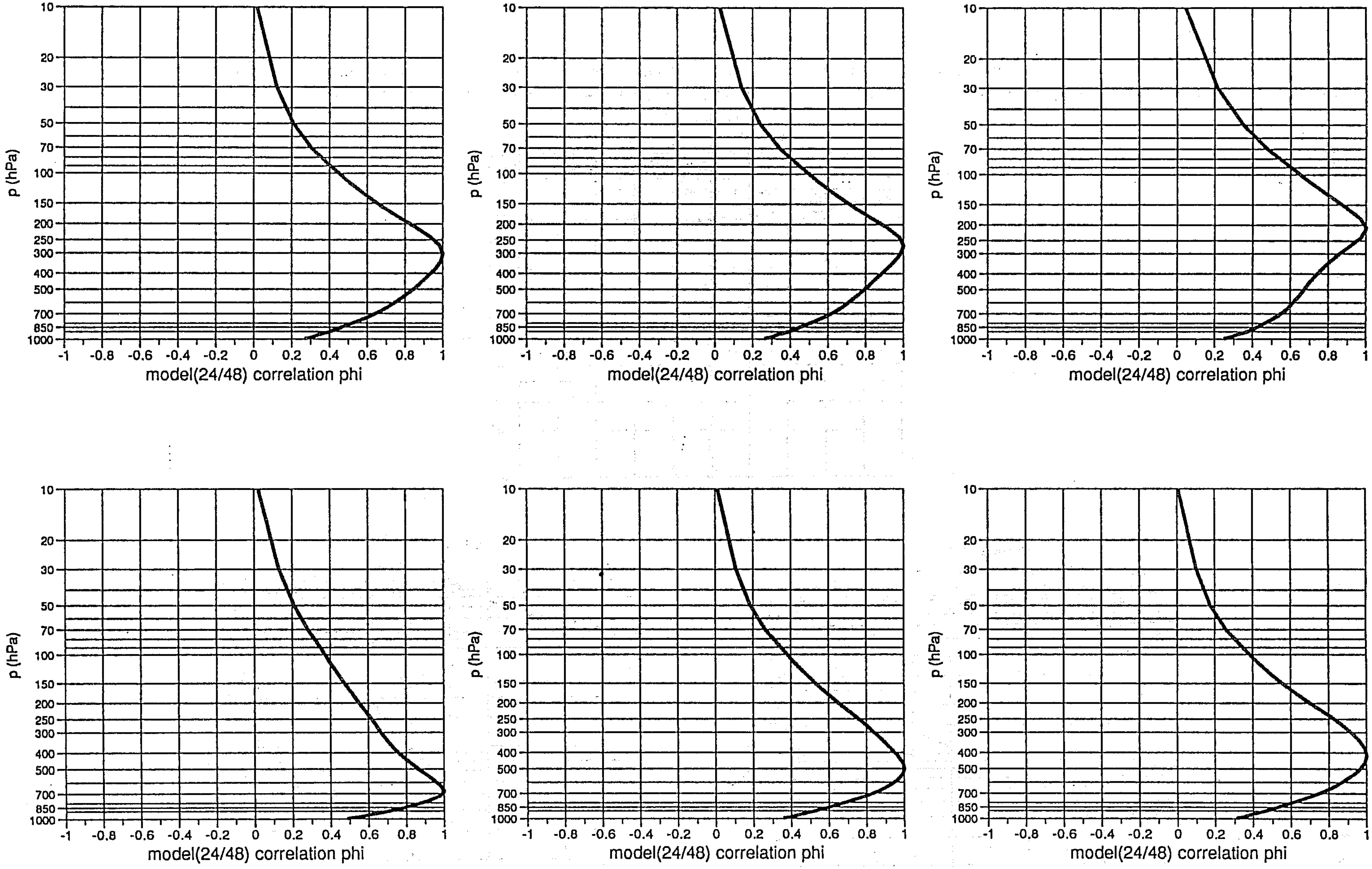

Fig. 11d) Vertical correlation of forecast errors as computed using NMC's approach: geopotential.

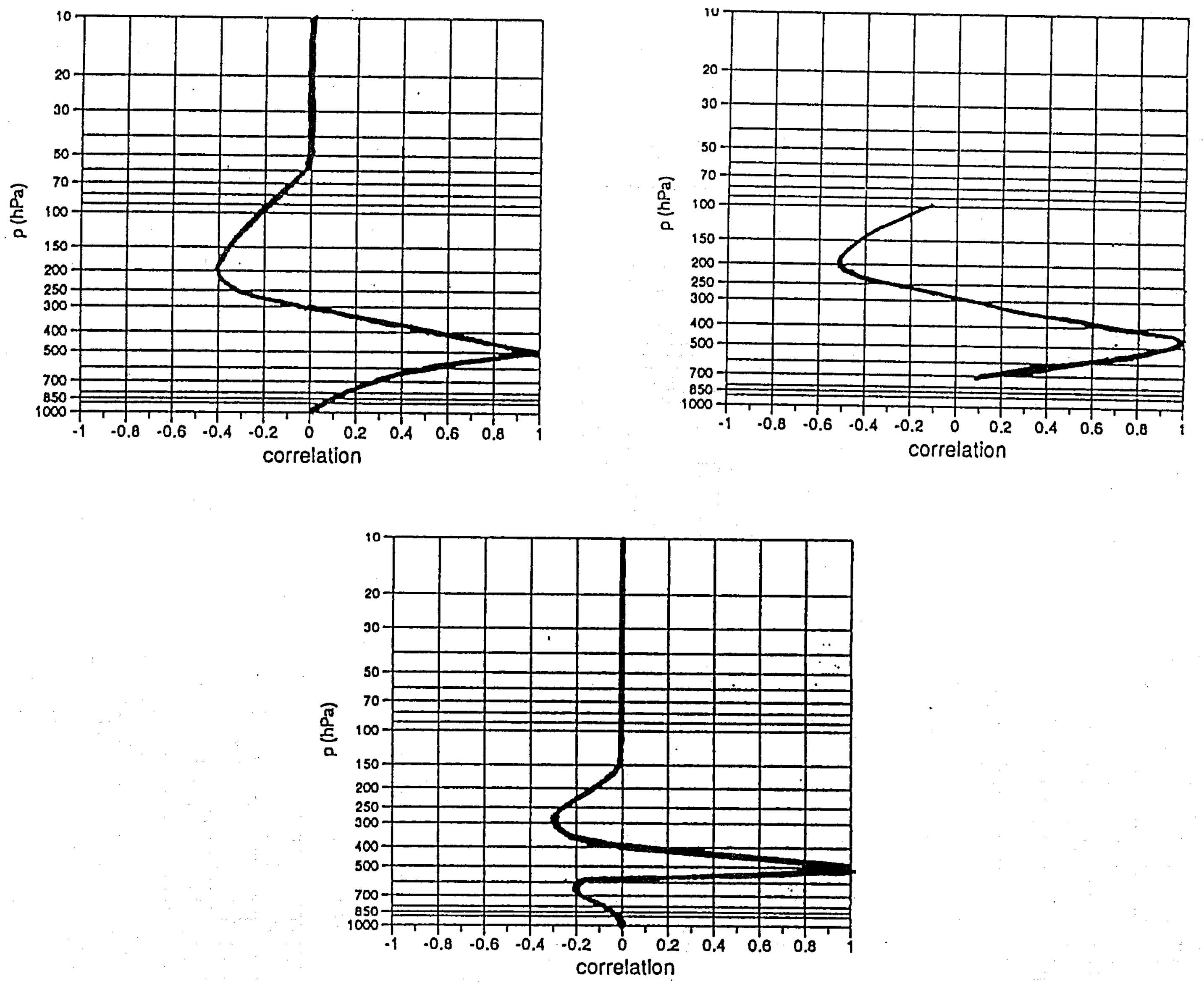

Fig. 12 Vertical correlations of temperature forecast errors, top left: derived from statistics background - radiosonde and used in 1D-Var, top right: derived from statistics background - radiosonde temperature thickness, bottom: currently used in OI.

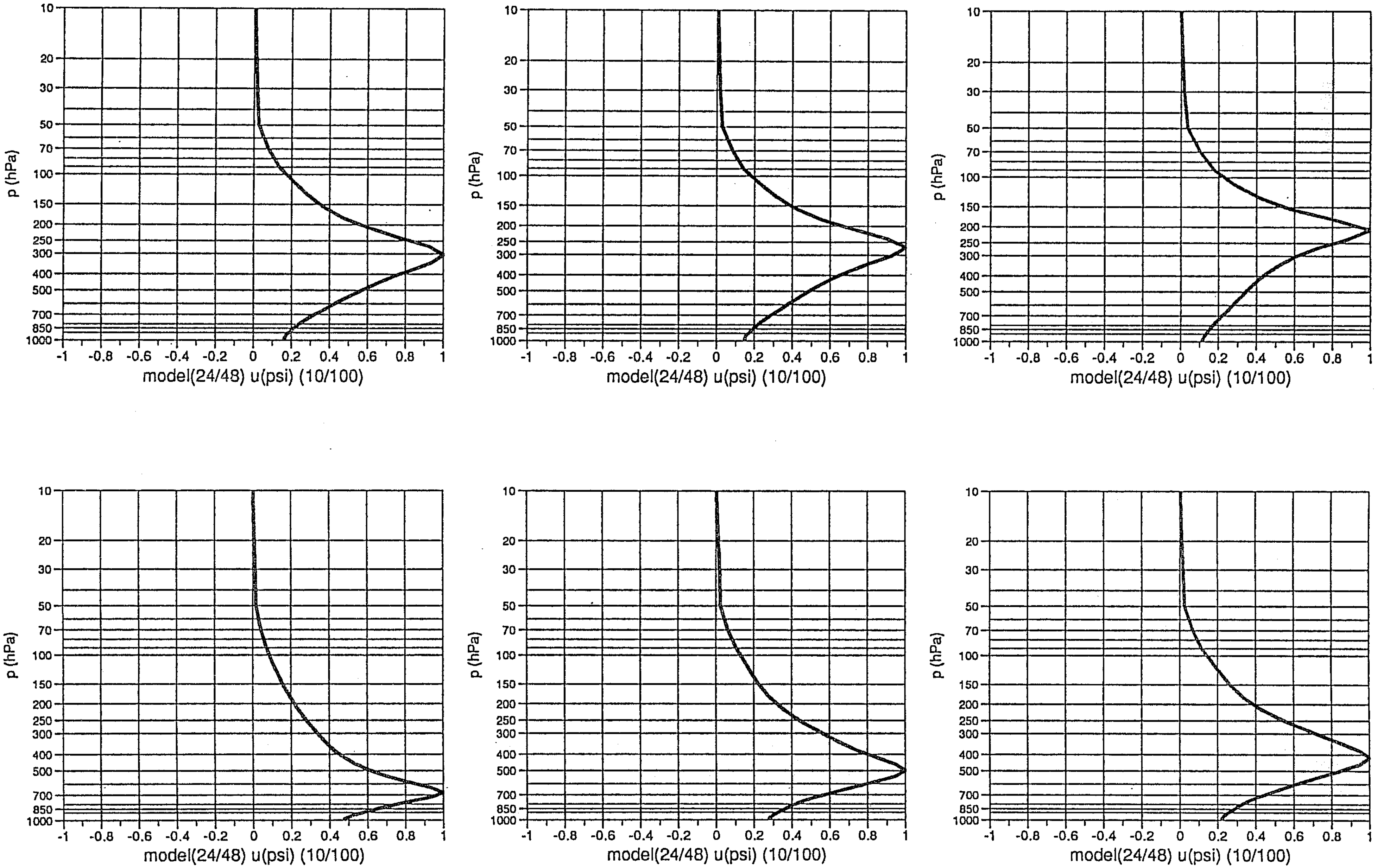

Fig. 13a) Vertical correlations for the synoptic scales (wavenumber 10 to 100) of rotational part of the wind.

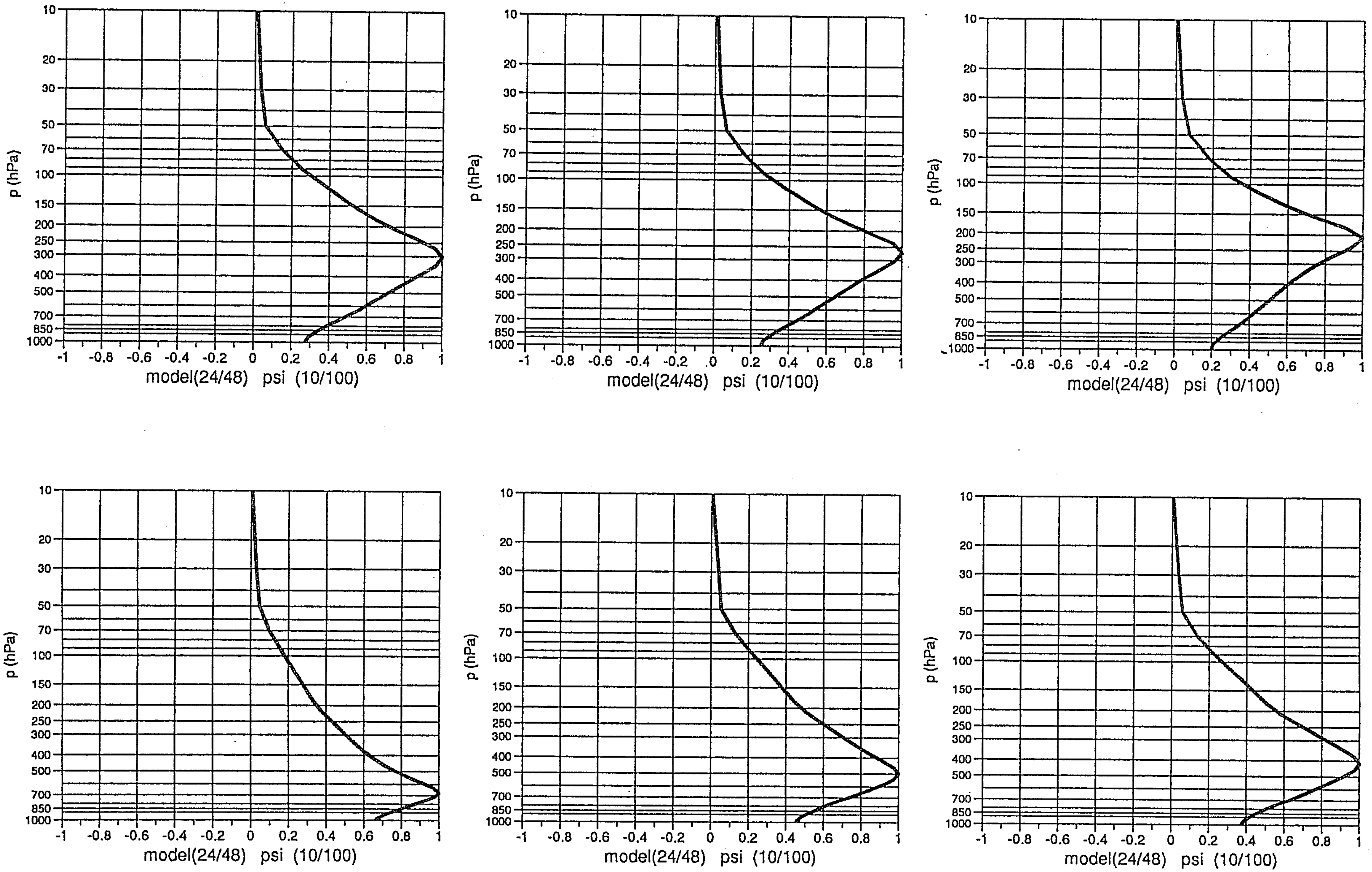

Fig. 13b) Vertical correlations for the synoptic scales (wavenumber 10 to 100) of streamfunction.

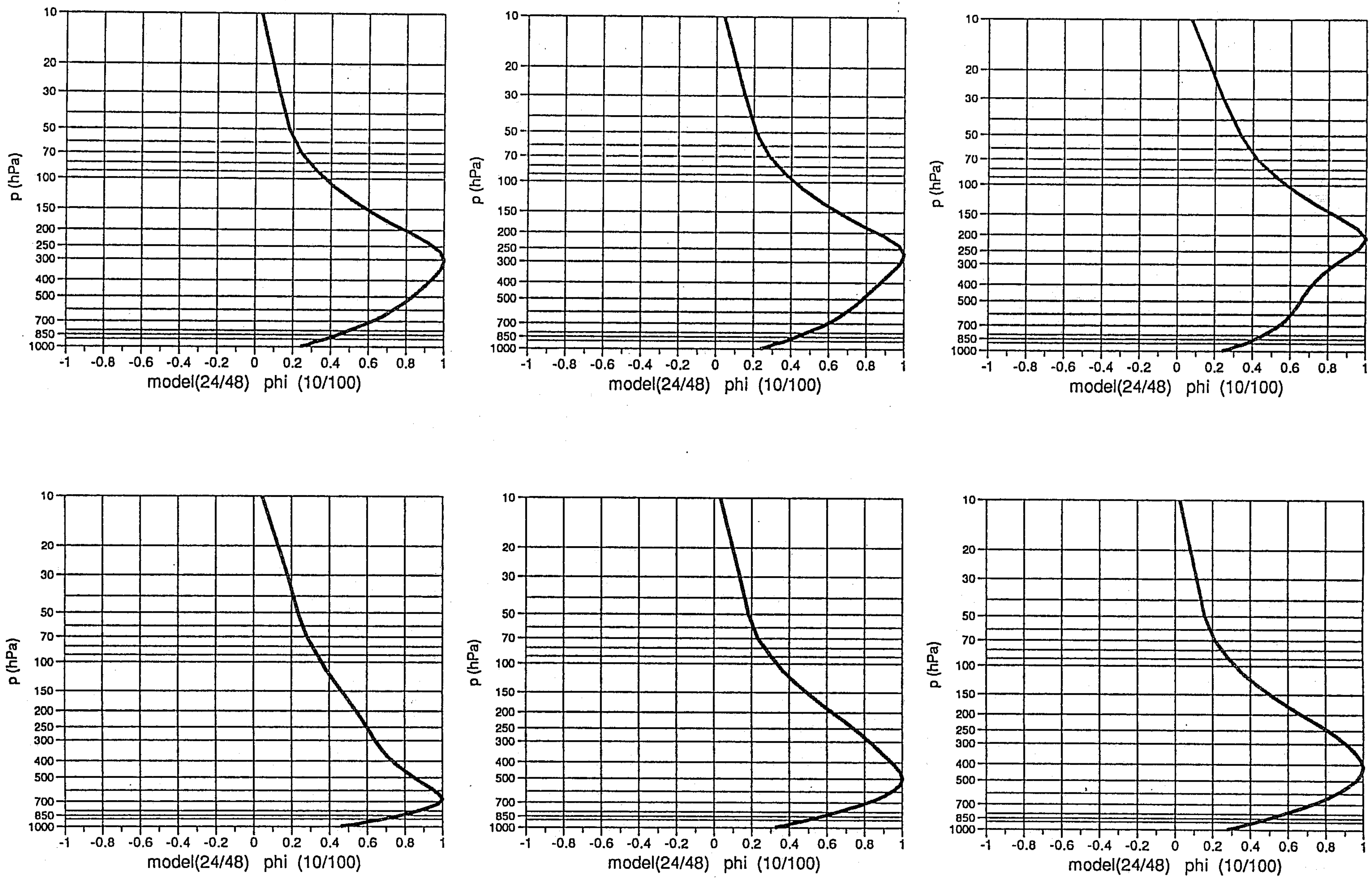

Fig. 13c) Vertical correlations for the synoptic scales (wavenumber 10 to 100) of geopotential.

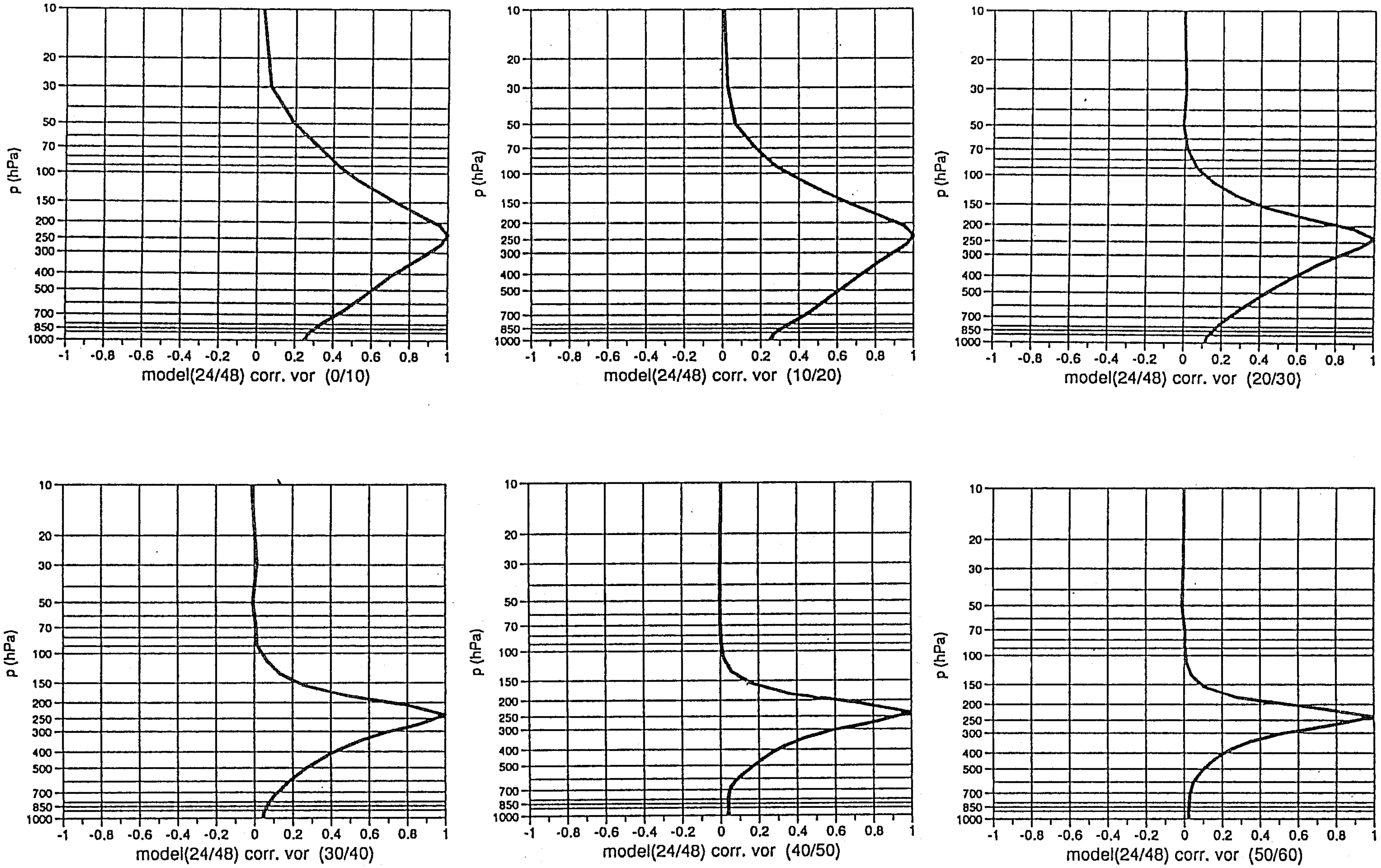

Fig. 14 Vertical correlations of the rotational part of the wind averaged over the wavenumber ranges n = 0 to 10, n = 10 to 20, n = 20 to 30, n = 30 to 40, n = 40 to 50 and n = 50 to 60.

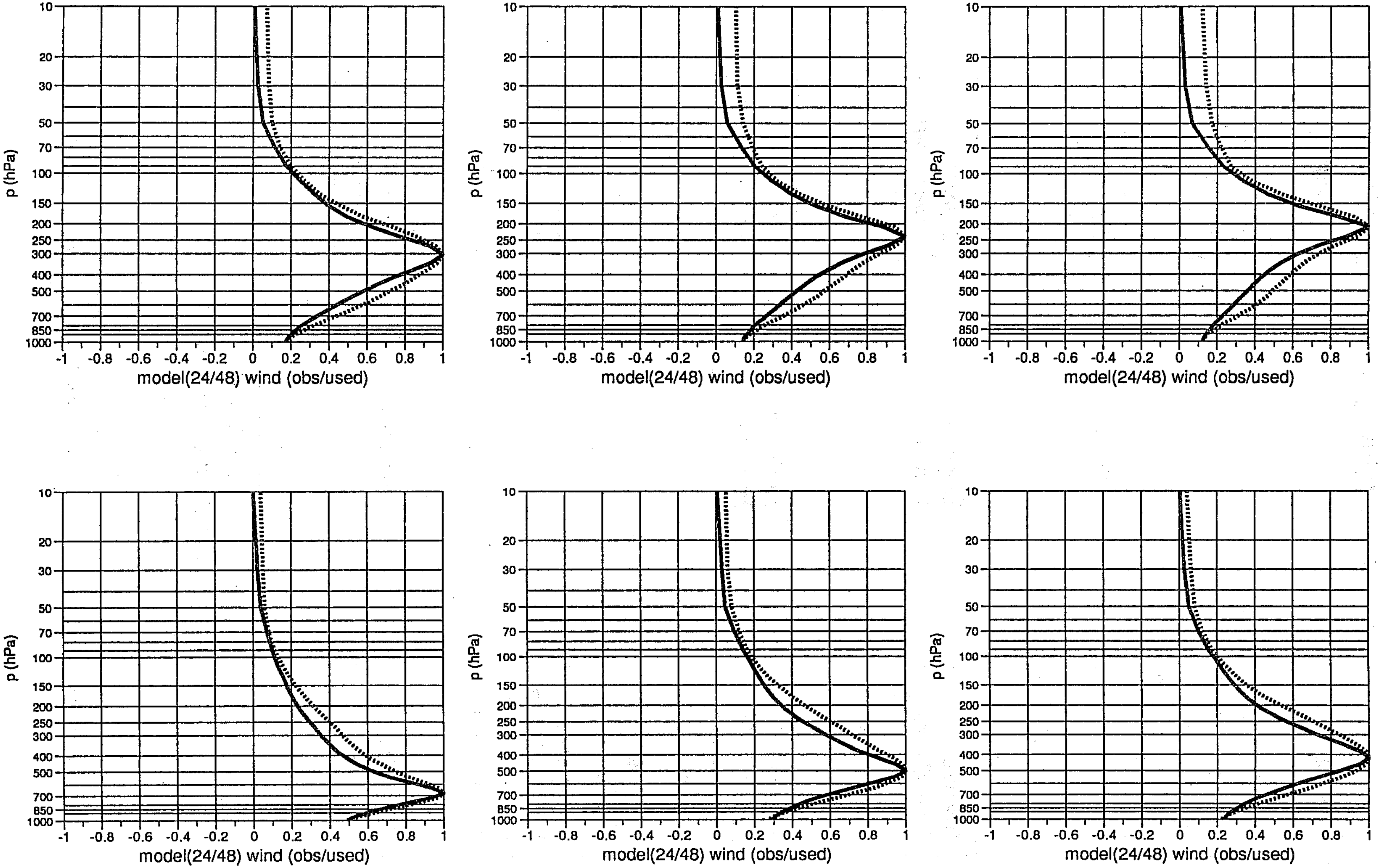

Fig. 15 The dotted lines show wind correlations as obtained from T, p_s correlations and the hydrostatic and geostrophic assumptions. The parametric spectrum displayed in Fig. 10 has been used. The solid lines show the directly observed wind correlations.

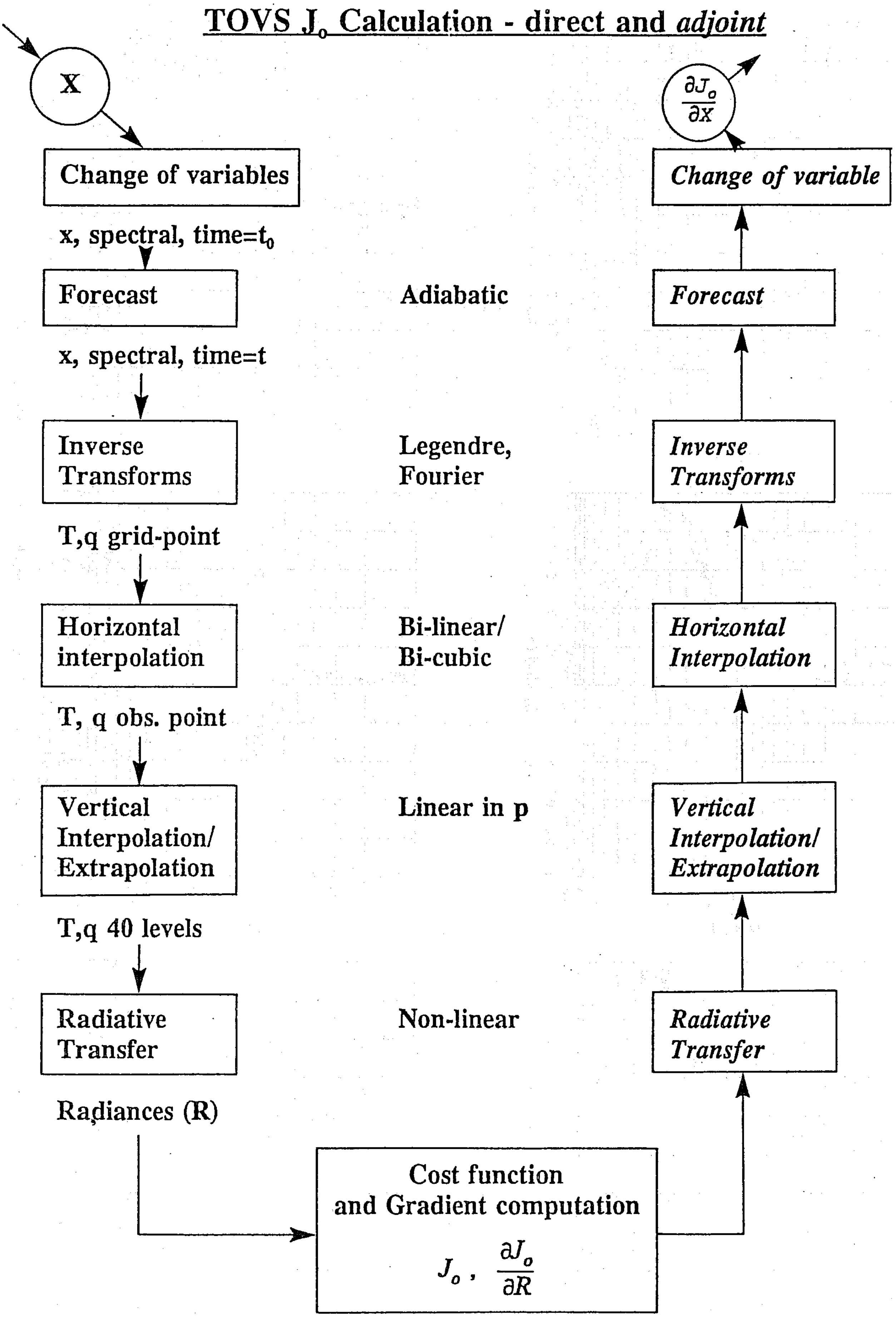

Fig. 16 Schematic representation of the cloud cleared radiances observation term computations.

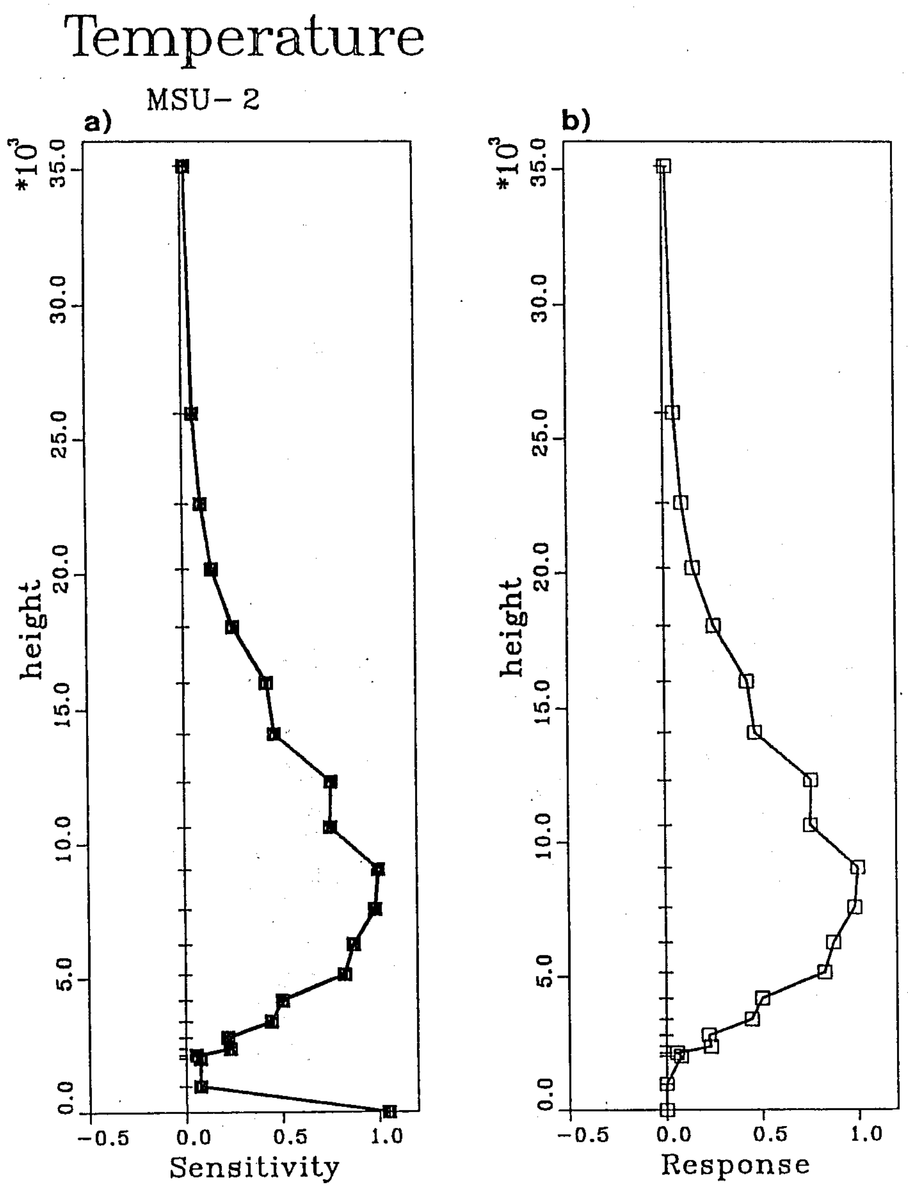

Fig. 17 a) MSU2 weight function
b) analysis increments using a single MSU2 channel and no vertical correlations of the background errors

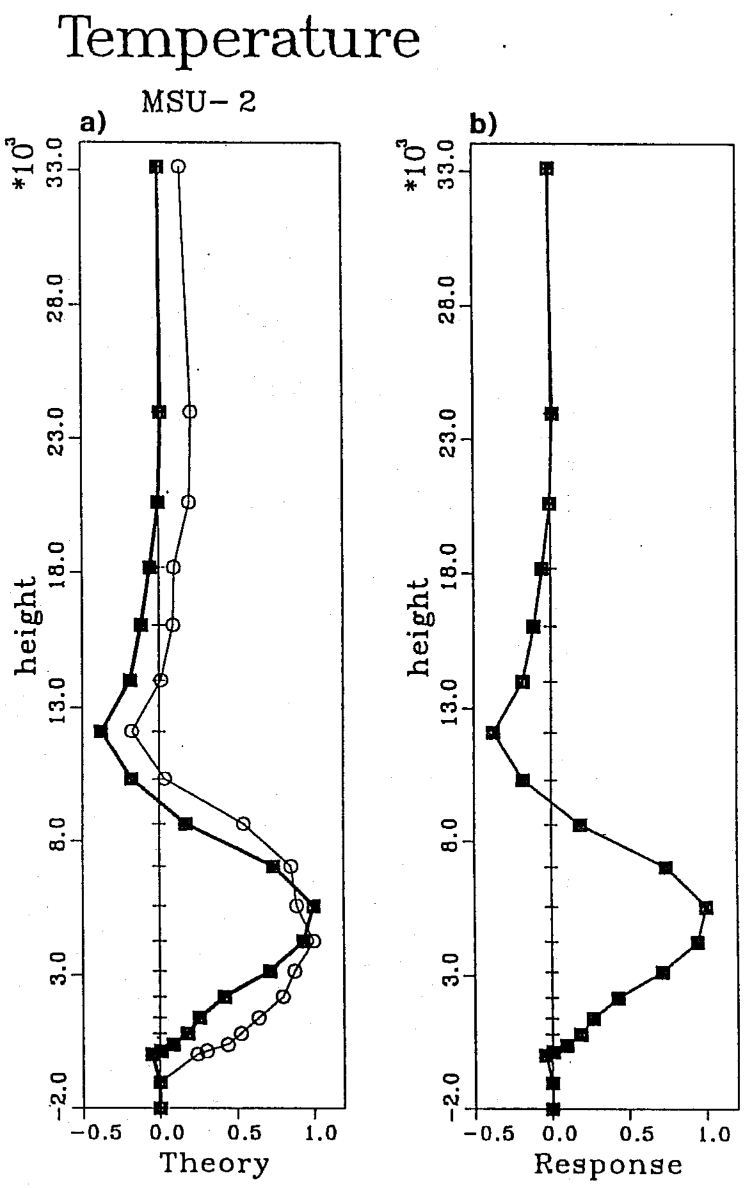

Fig. 18 a) theoretical analysis increment (thick line), 1D-Var increment (thin line), b) 3D-Var analysis increment. The agreement validates the 3D-Var code with respect to the use of radiances.

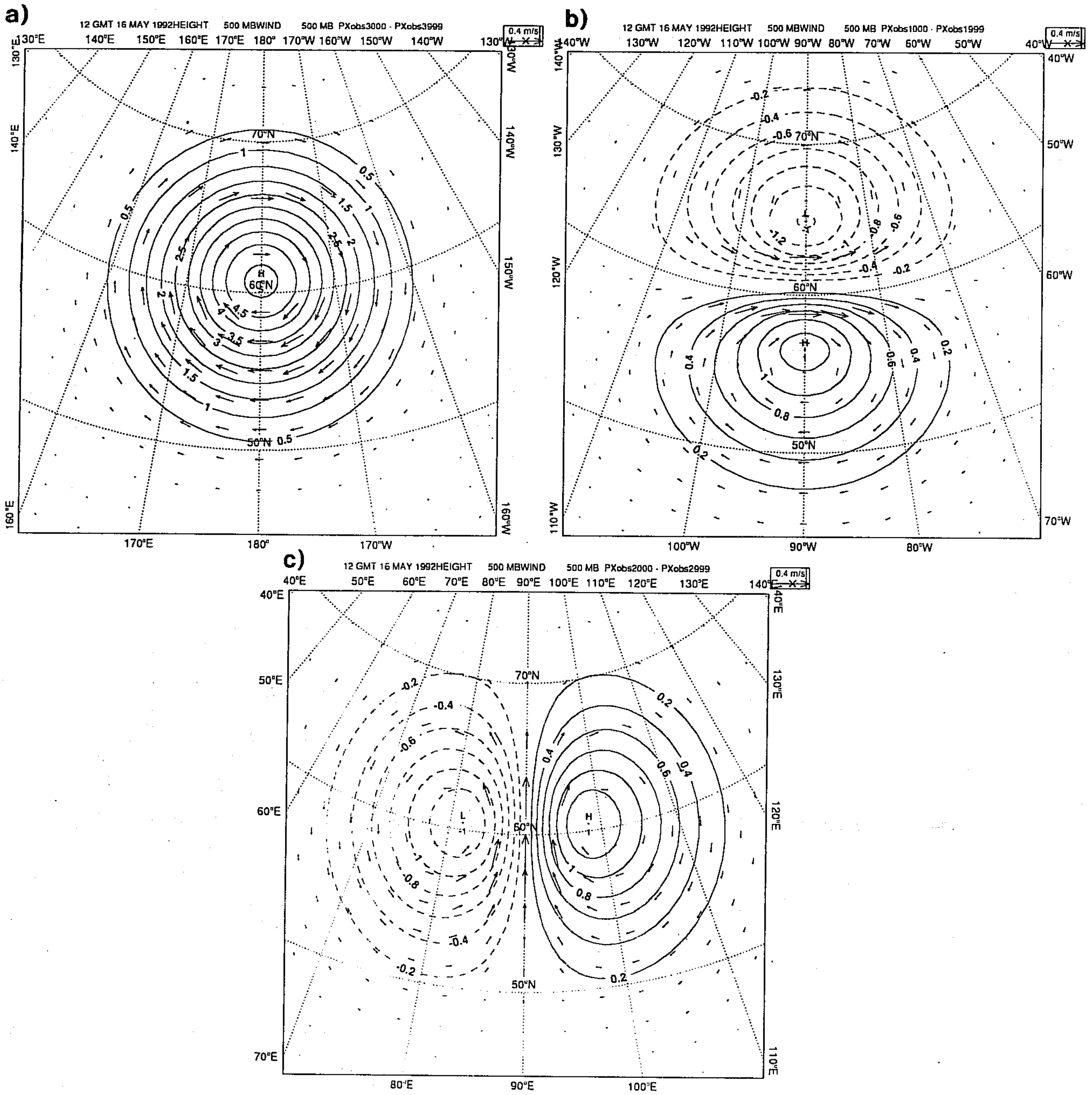

Fig. 19 Response of the 3D-Var analysis at 500 hPa to an isolated observation at 60°N. Hough balance ε = 0.1. a) height, observation, b) zonal wind observation, c) meridional wind observation. Contours are of height increments and the arrow indicates the vector wind increment.

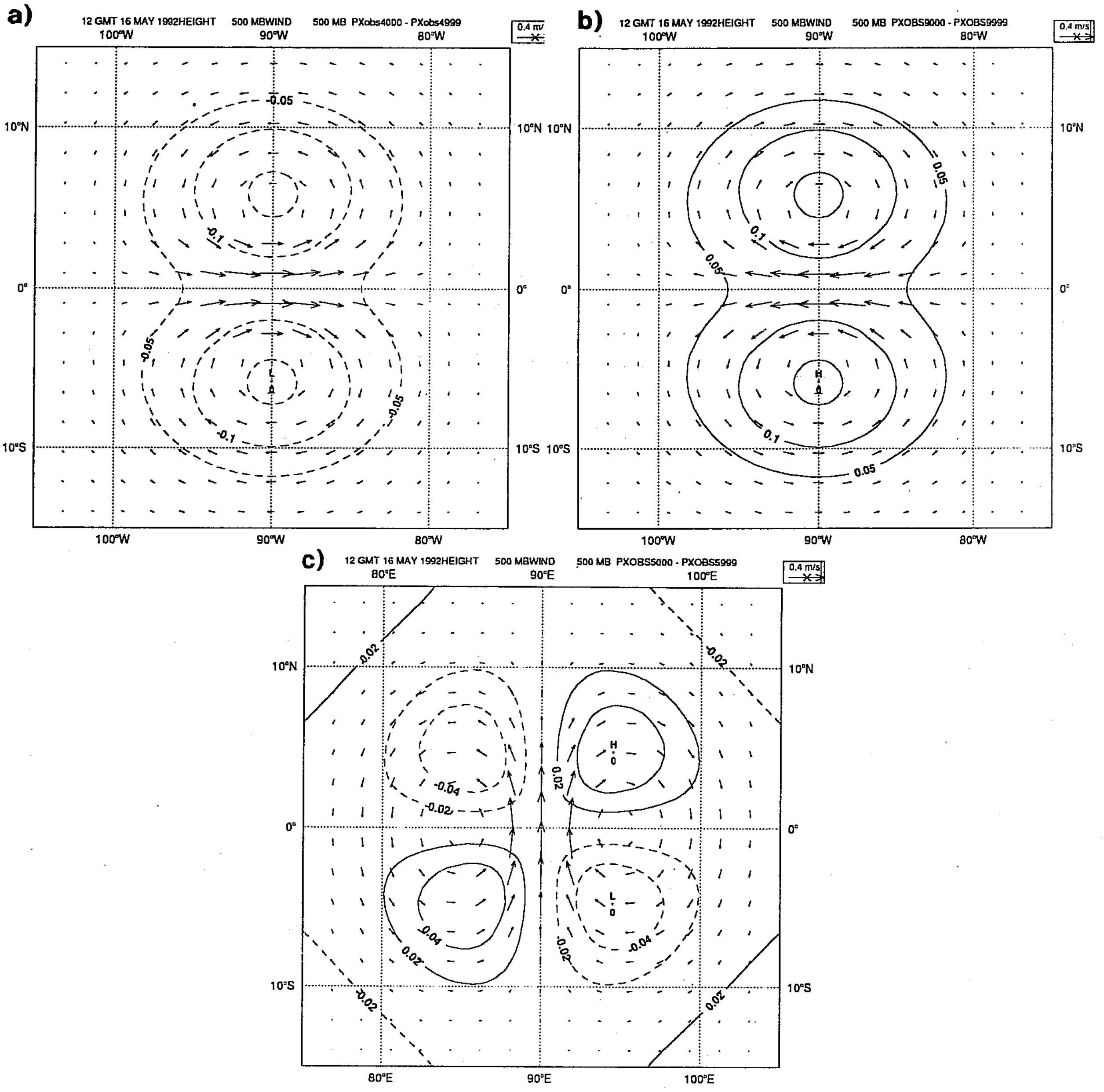

Fig. 20 Response of the 3D-Var analysis at 500 hPa to an isolated observation at 0°N. Hough balance ε = 0.1. a) positive zonal wind, b) negative zonal wind, c) southerly meridional wind. Contours are of height and the arrows indicate vector wind.

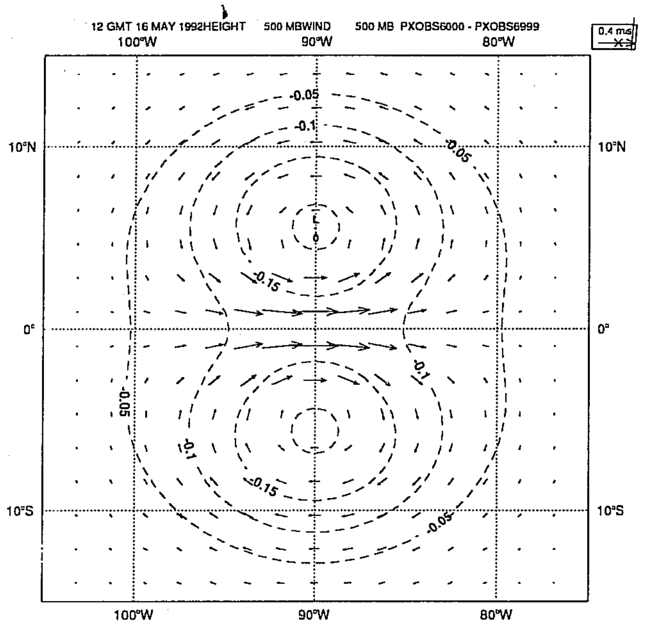

Fig. 21 As Fig. 20(a) but for linear balance.

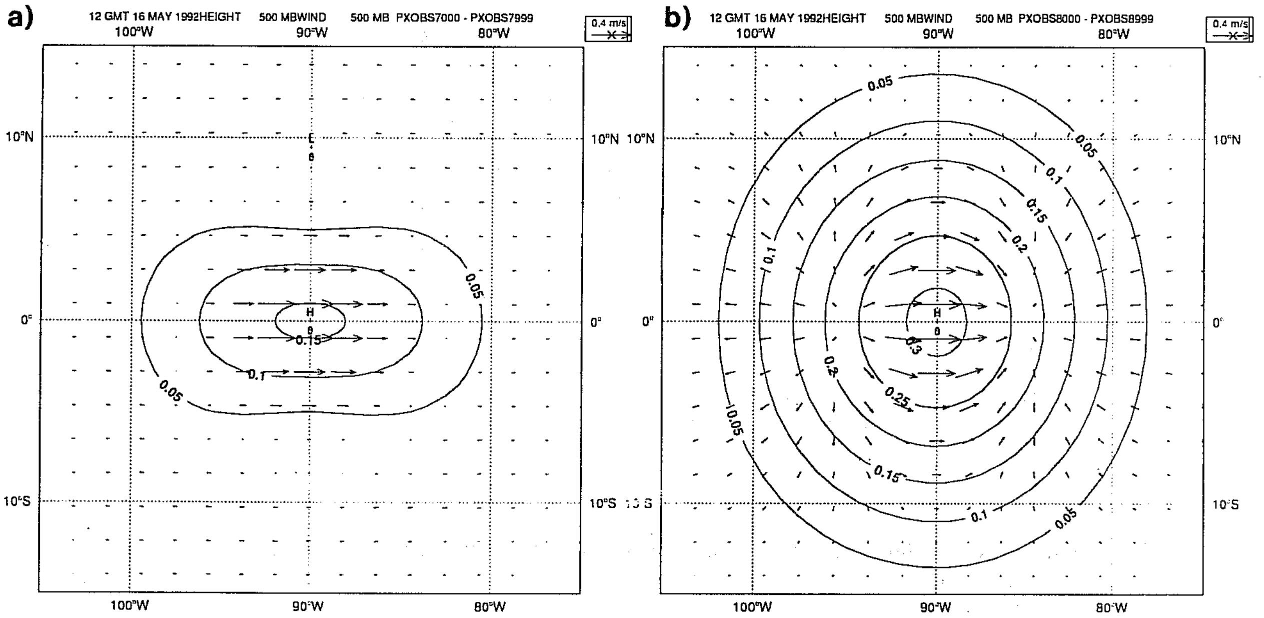

Fig. 22 As Fig. 20(a) but for a) ε = 0.5 and b) ε = 0.9.

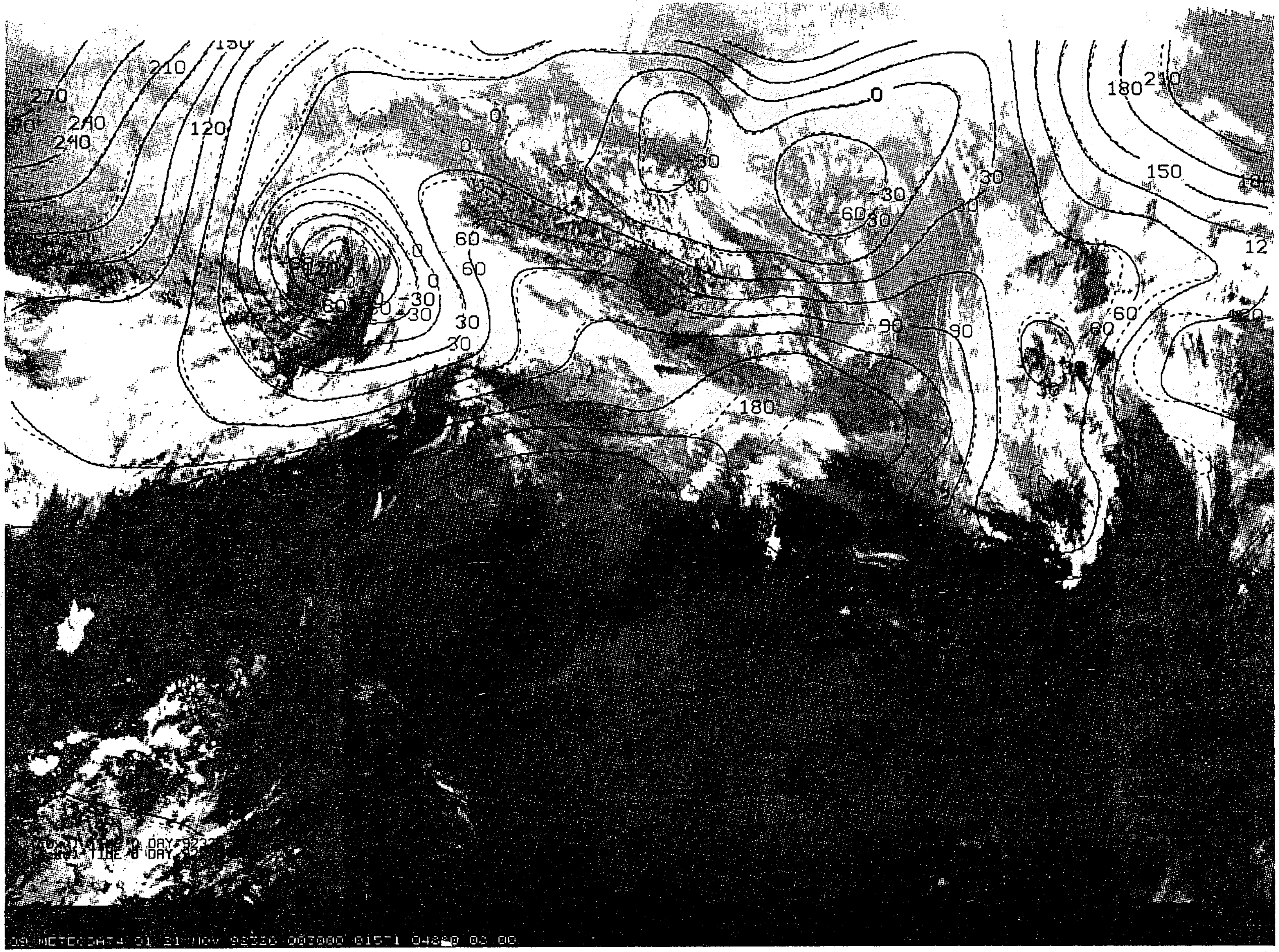

Fig. 23a) 1000 hPa height after 10 days of assimilation solid 3D-Var, dashed OI 1D-Var superimposed to the METEOSAT 11 micron imagery, Atlantic Ocean.

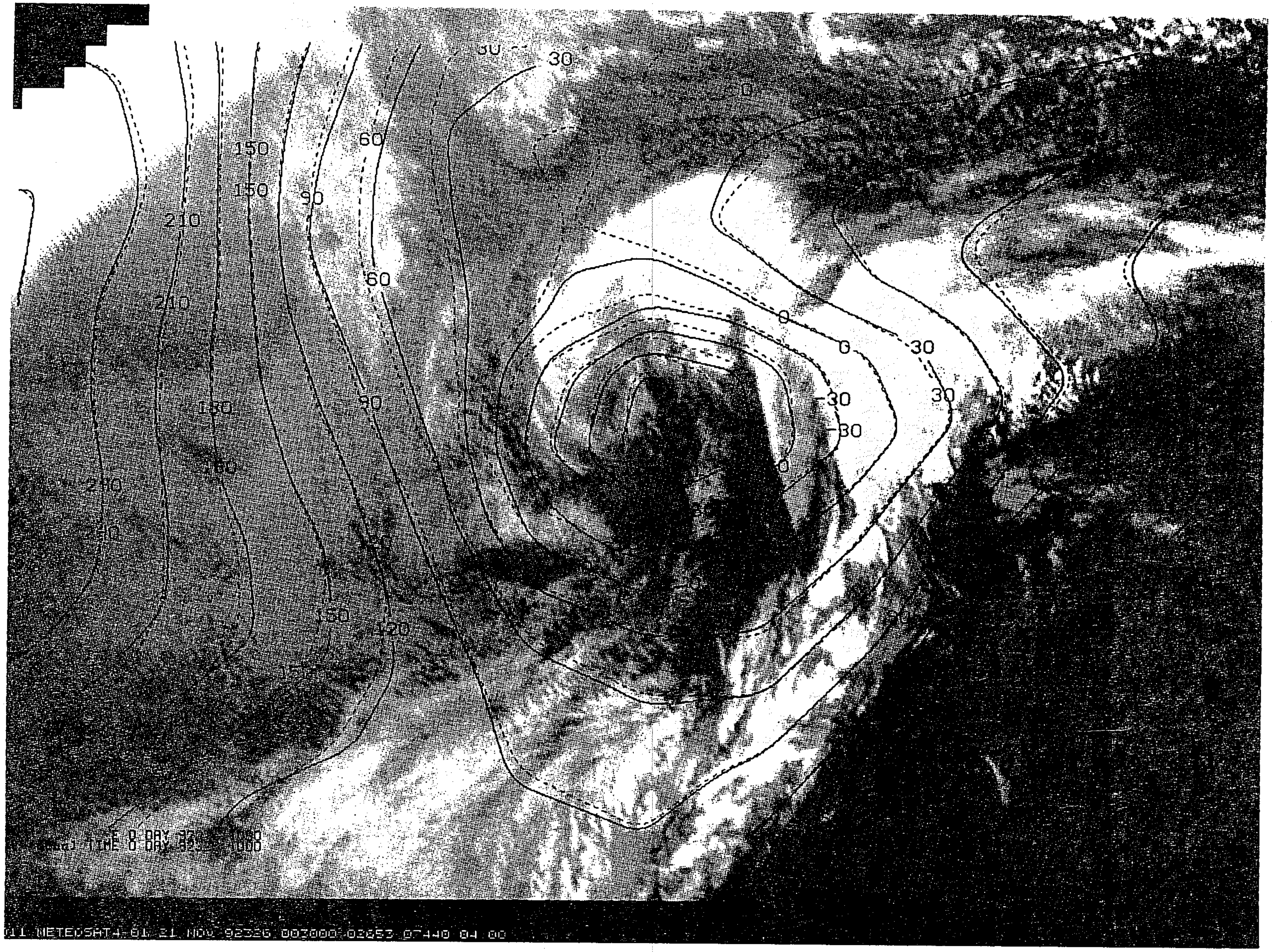

Fig. 23b) 1000 hPa height after 10 days of assimilation solid 3D-Var, dashed OI 1D-Var superimposed to the METEOSAT 11 micron imagery, zoom over the low.

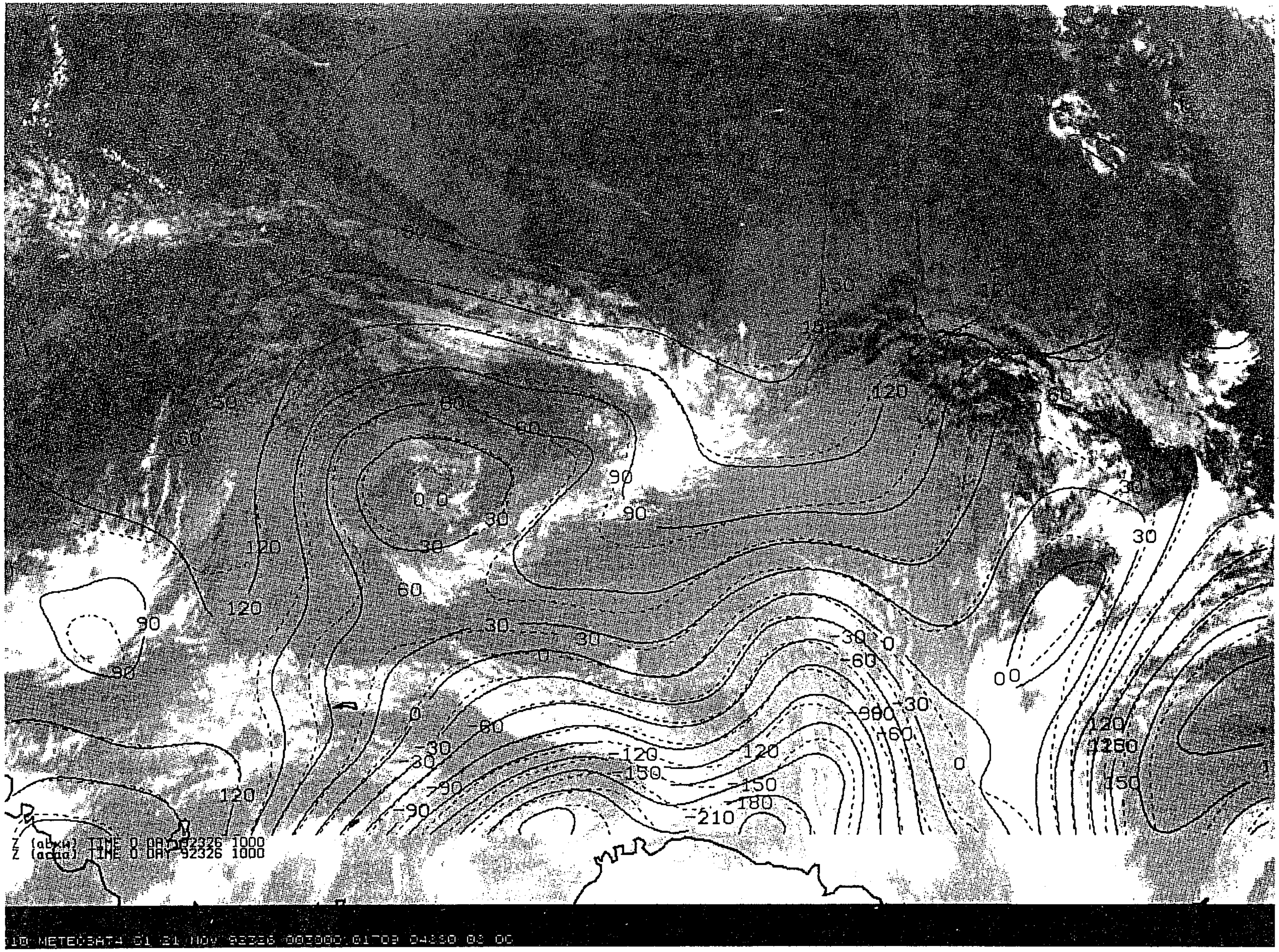
Z (abxw) TIME 0 DAY 92326 1000
Z (acqa) TIME 0 DAY 92326 1000
10 METEOSAT4 01 21 NOV 92326 003000 01703 04330 02 00
120
90
60
30
0
-30
-60
-90
-120
-150
-180
-210

Fig. 23c) 1000 hPa height after 10 days of assimilation solid 3D-Var, dashed OI 1D-Var superimposed to the METEOSAT 11 micron imagery, South Pacific.

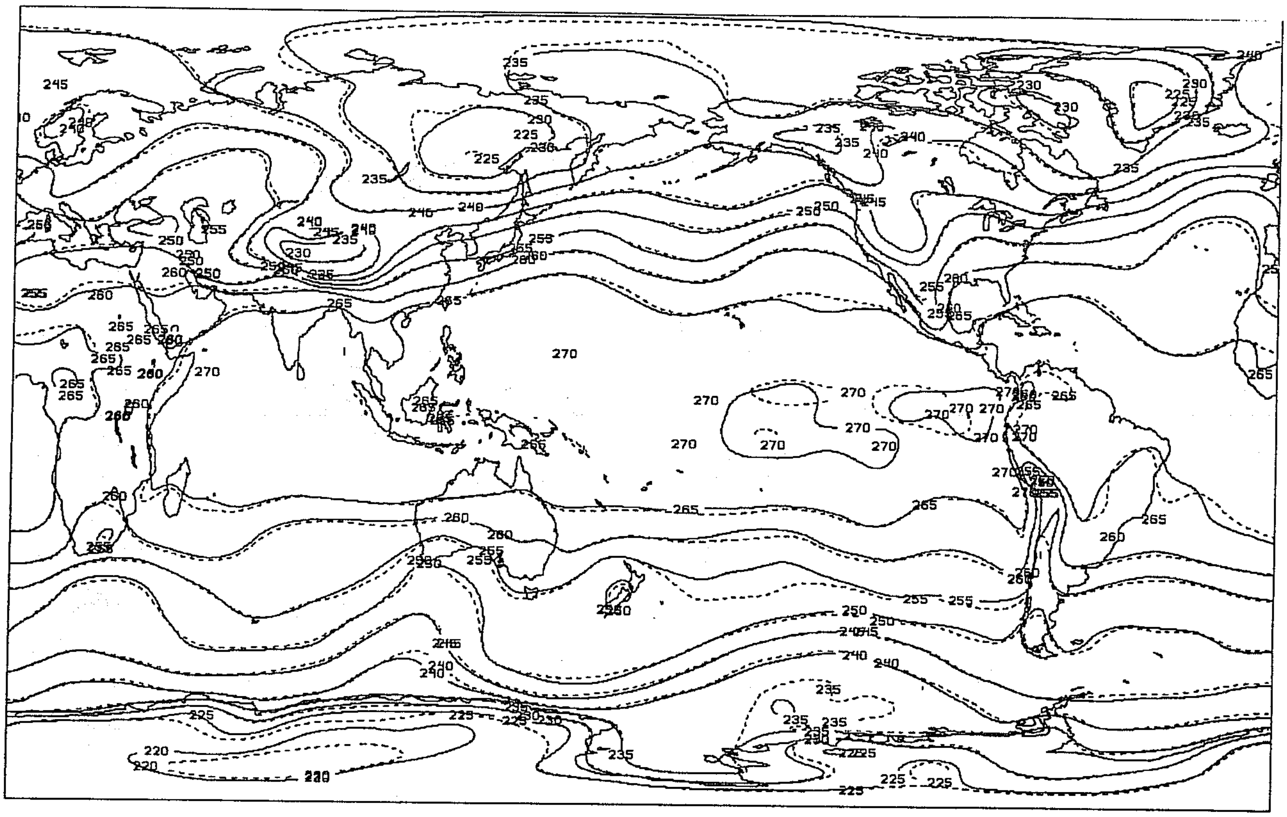

Fig. 24a) Average of the analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var. Model level 11 temperature (~ 500 hPa).

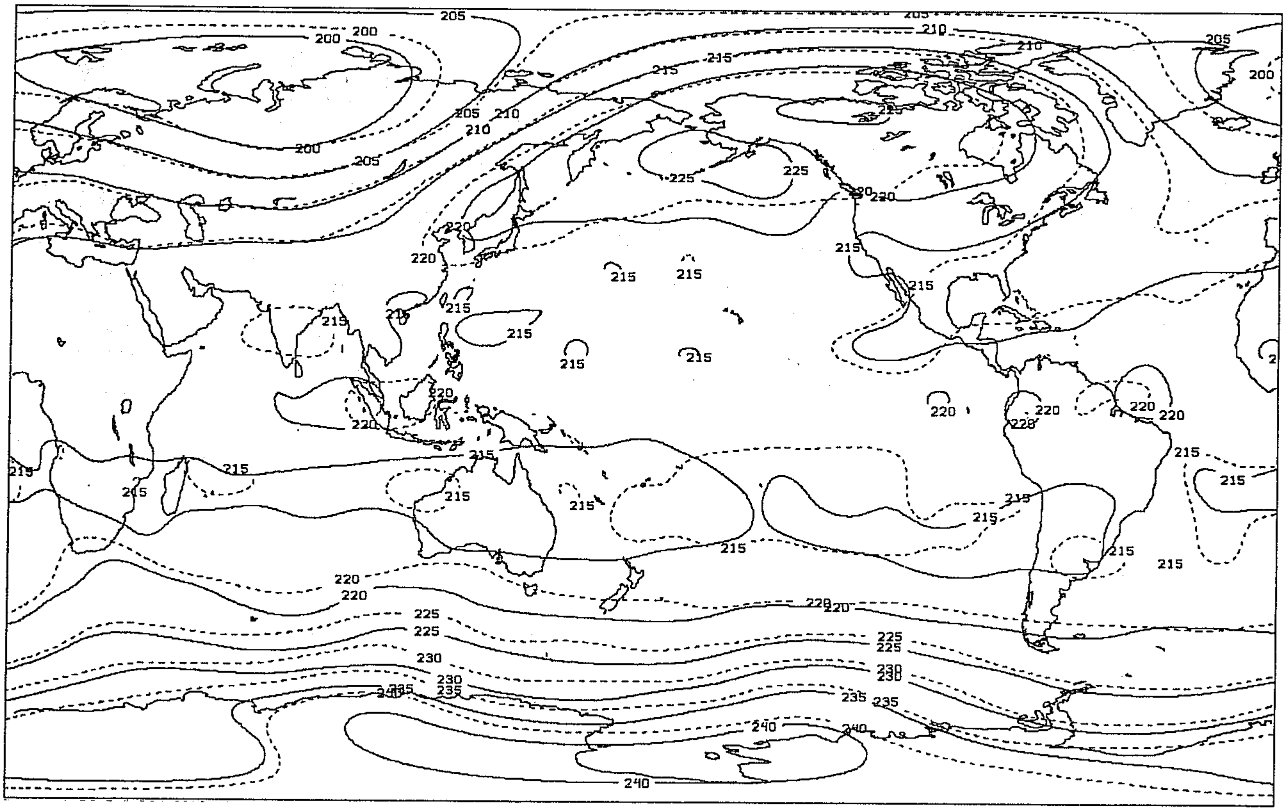

Fig. 24b) Average of the analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var. Model level 2 temperature (50 hPa).

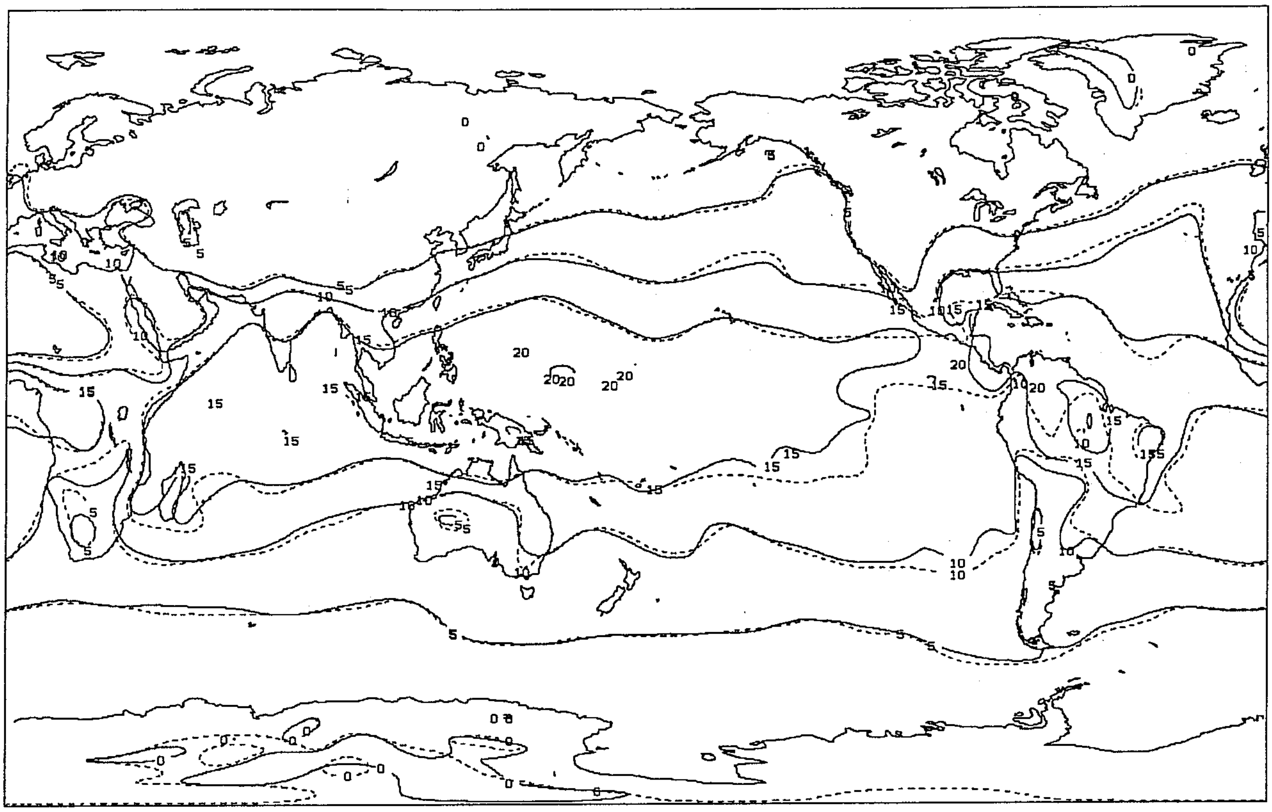

Fig. 24c) Average of the analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var. Model level 19 humidity (surface).

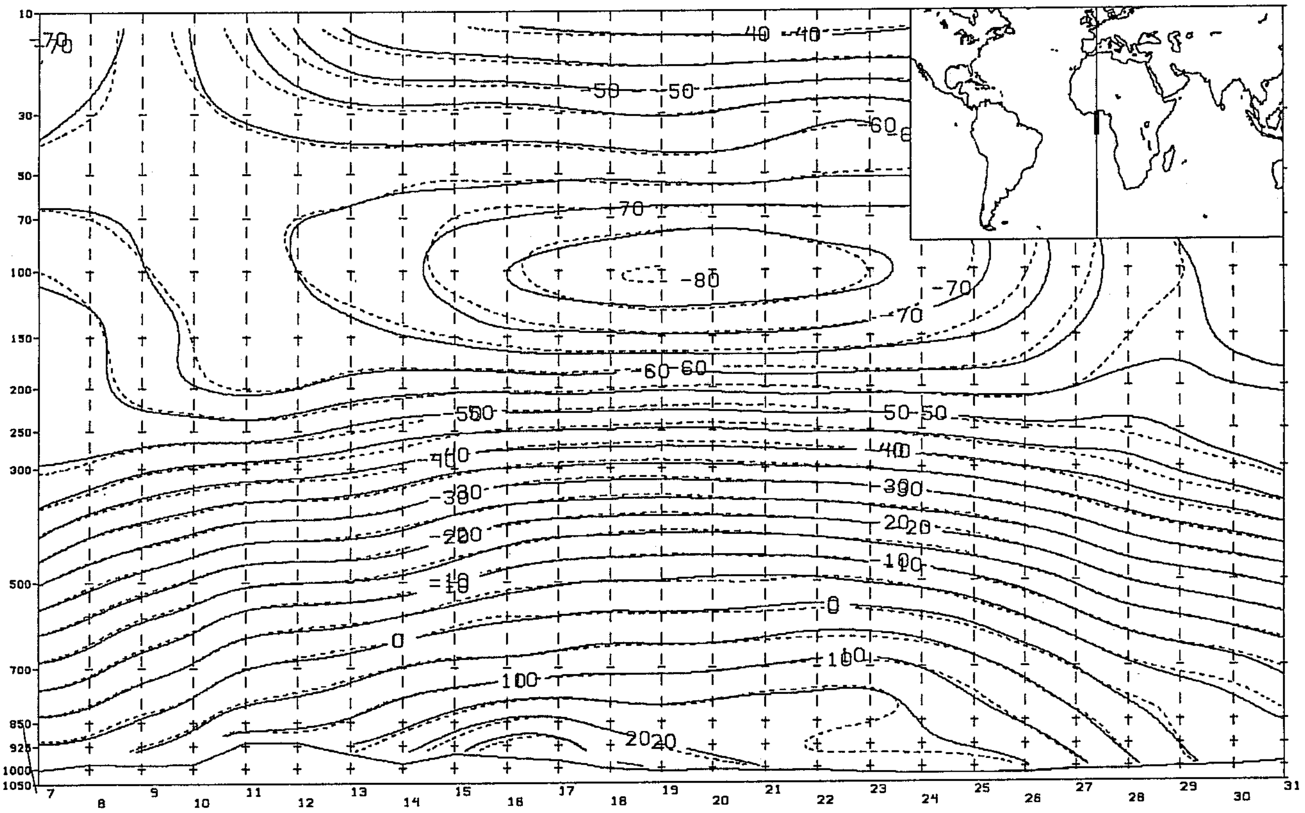

Fig. 25a) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var, temperature.

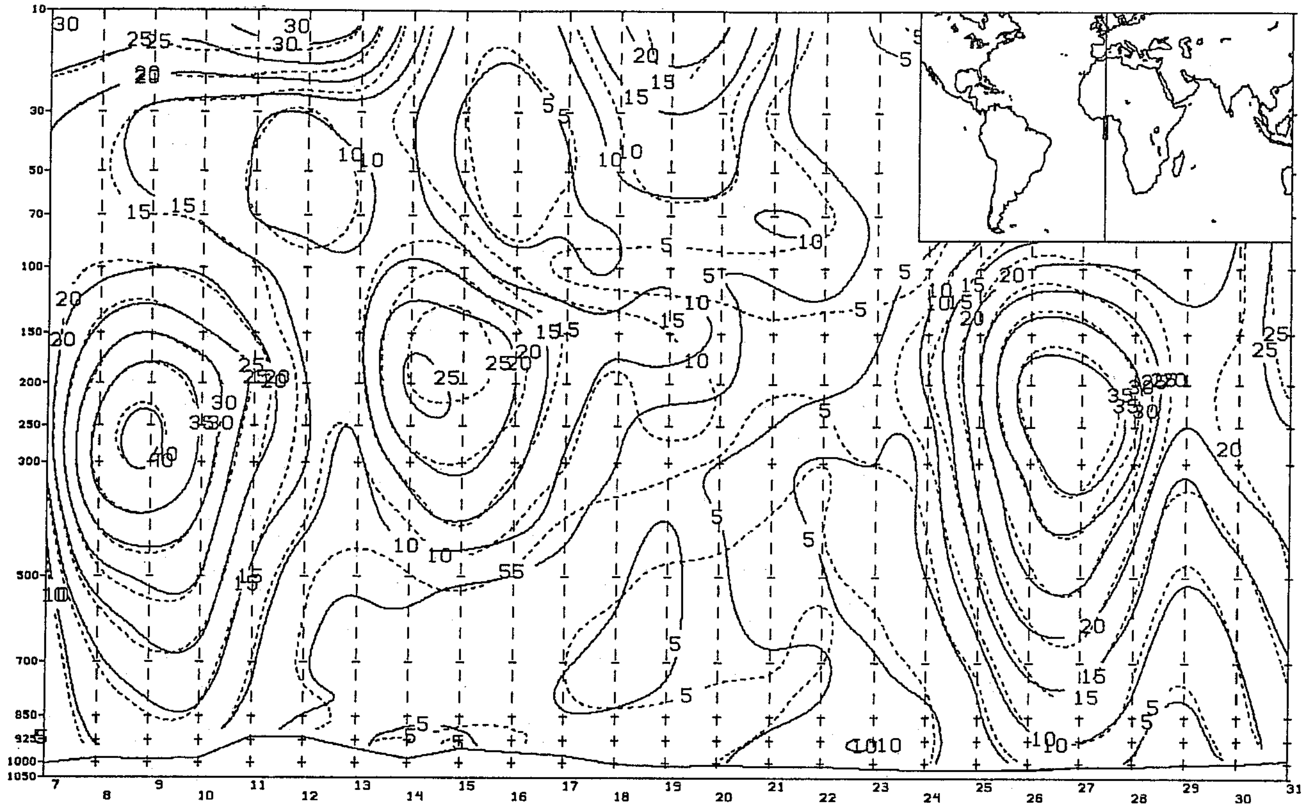

Fig. 25b) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var, wind.

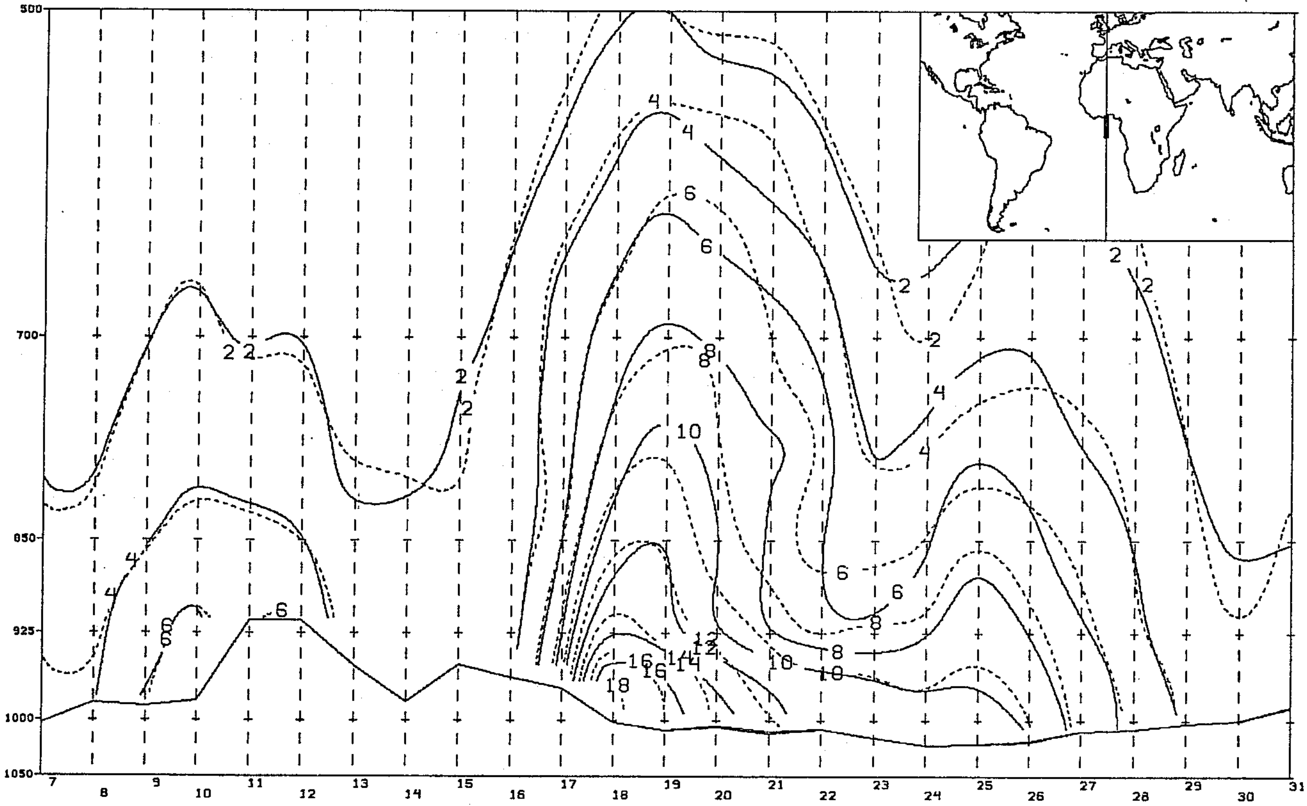

Fig. 25c) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed OI 1D-Var, humidity.

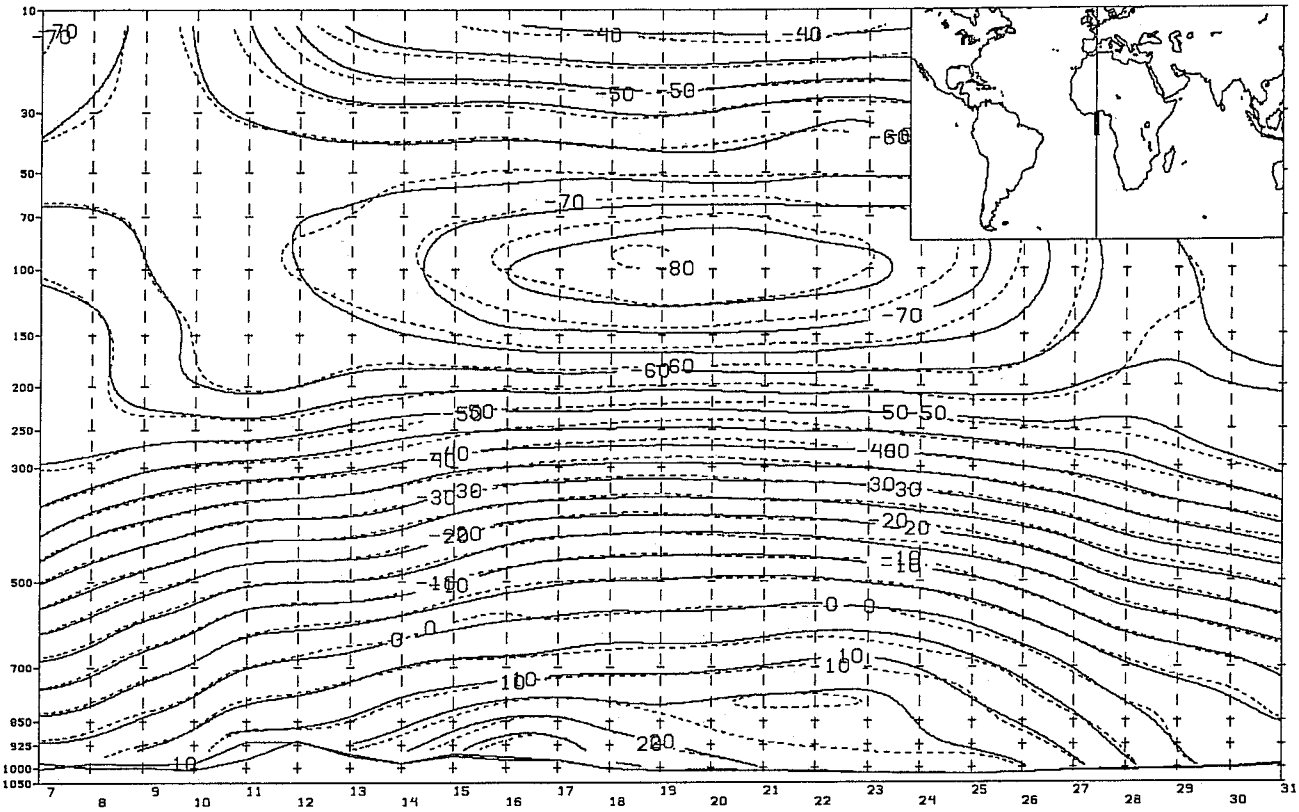

Fig. 26a) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed operational analyses, temperature.

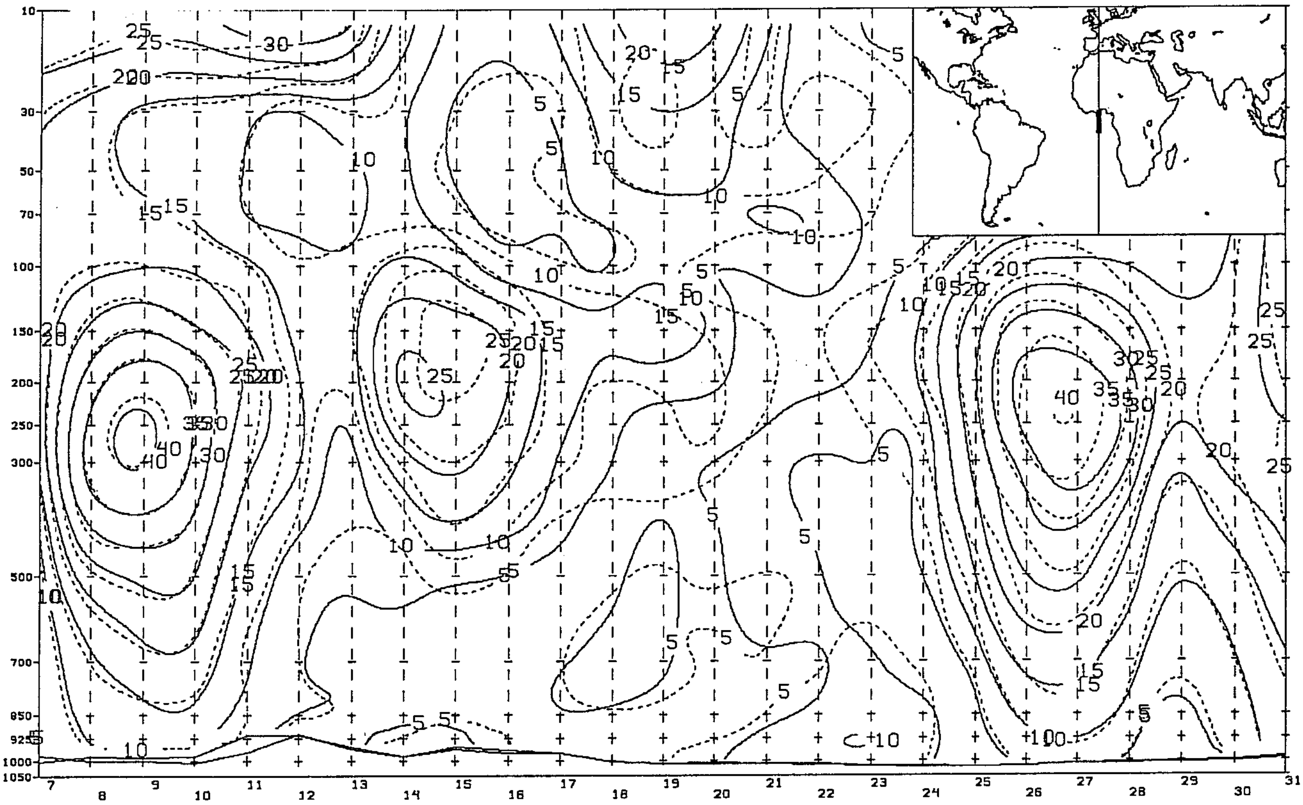

Fig. 26b) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed operational analyses, wind.

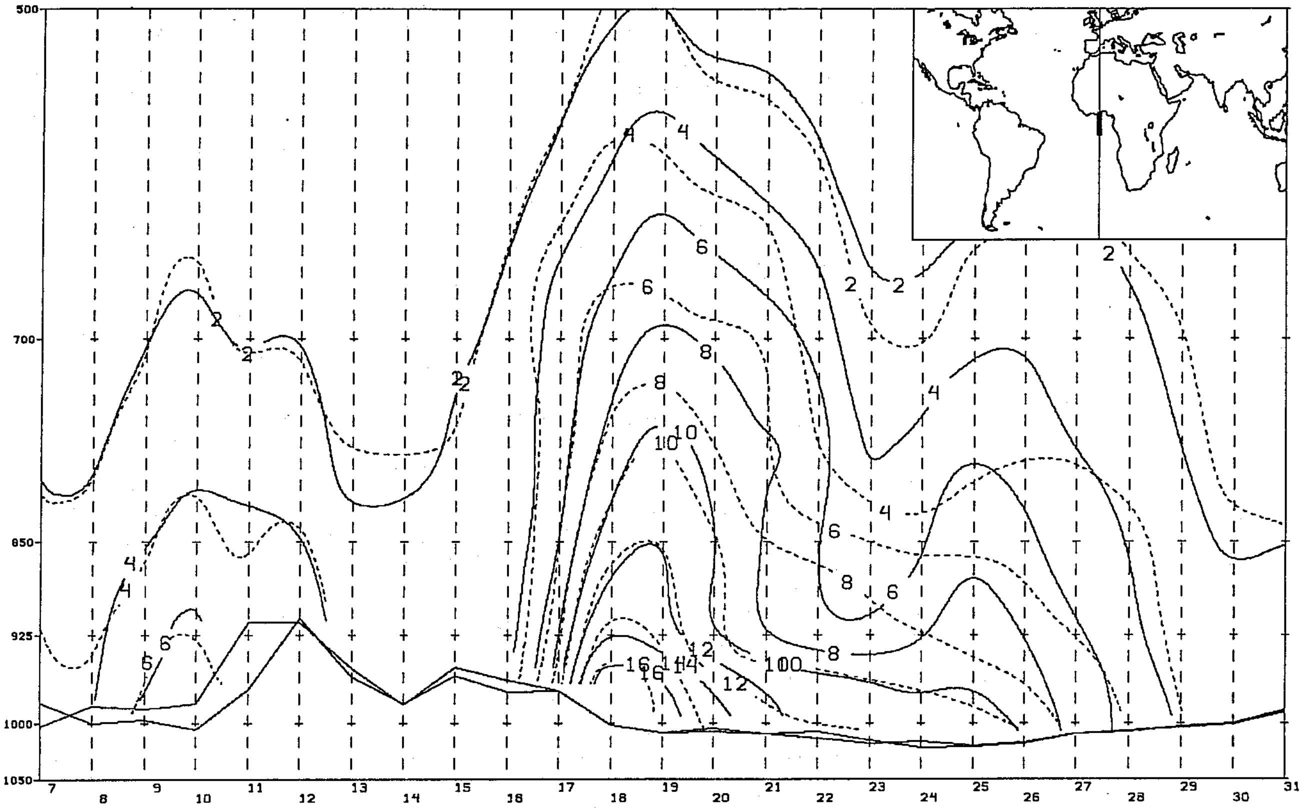

Fig. 26c) Cross section (meridian O) of the average of one week of analyses after 5 days of assimilation. Solid 3D-Var, dashed operational analyses, humidity.

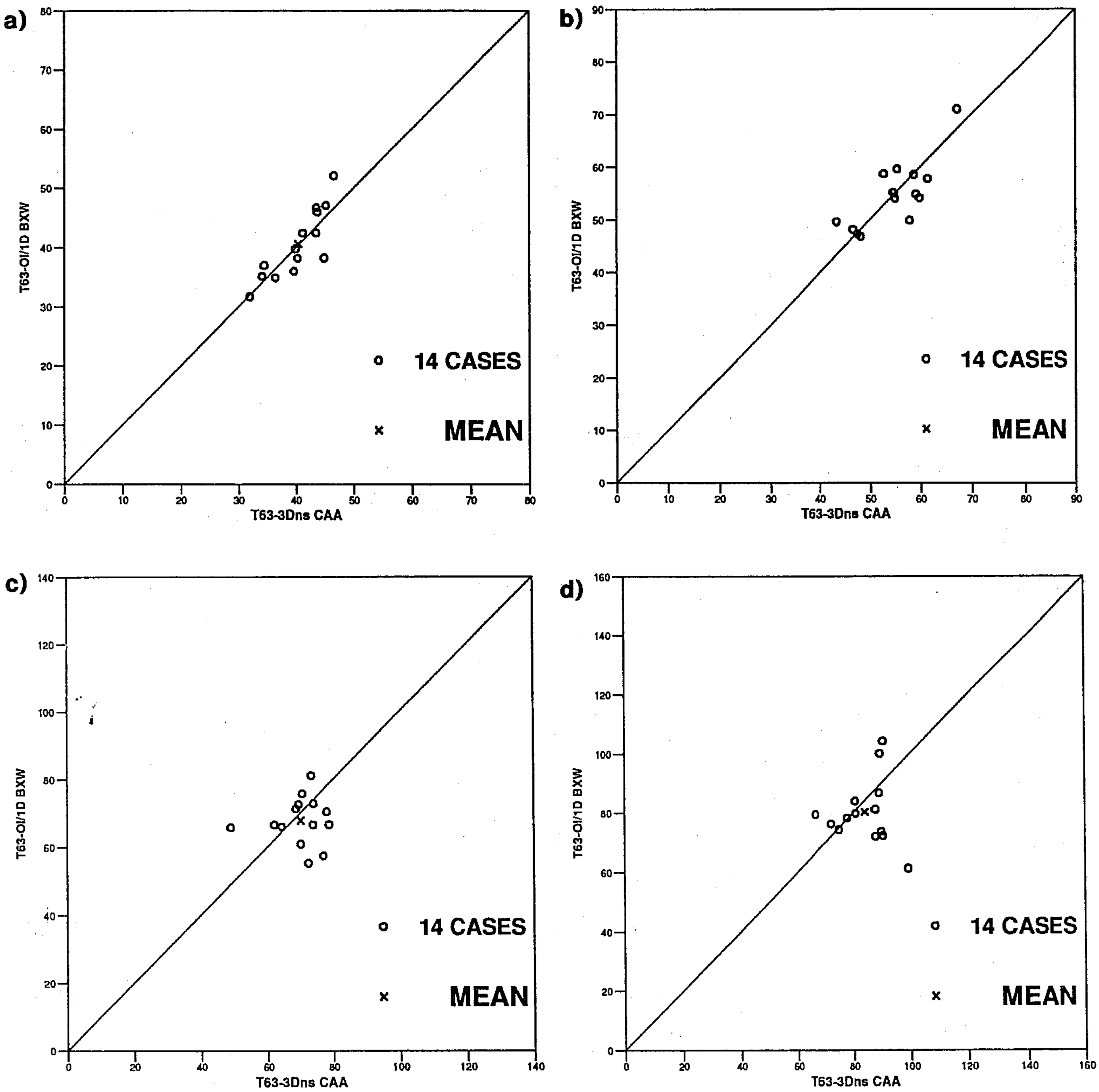

Fig. 27(A) Scatter diagram of the rms of the 500 hPa forecast error issued from 3D-Var (horizontal) and from OI/1D-Var (vertical) verified against the operational 500 hPa height analyses: Northern Hemisphere.
a) 3 day range, b) 4 day range, c) 5 day range, d) 6 day range.

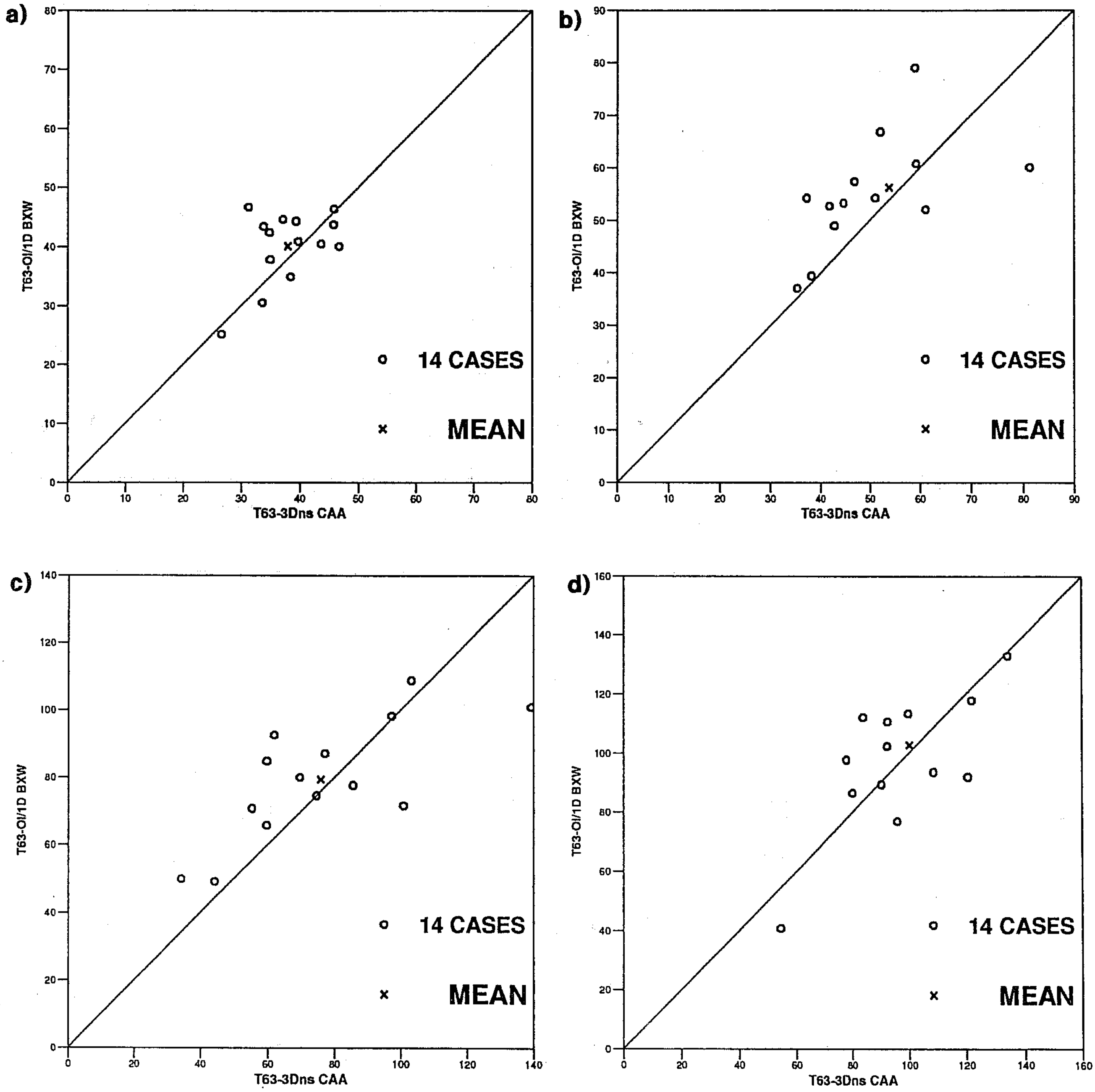

Fig. 27(B) Scatter diagram of the rms of the 500 hPa forecast error issued from 3D-Var (horizontal) and from OI/1D-Var (vertical) verified against the operational 500 hPa height analyses: Europe.
a) 3 day range, b) 4 day range, c) 5 day range, d) 6 day range.

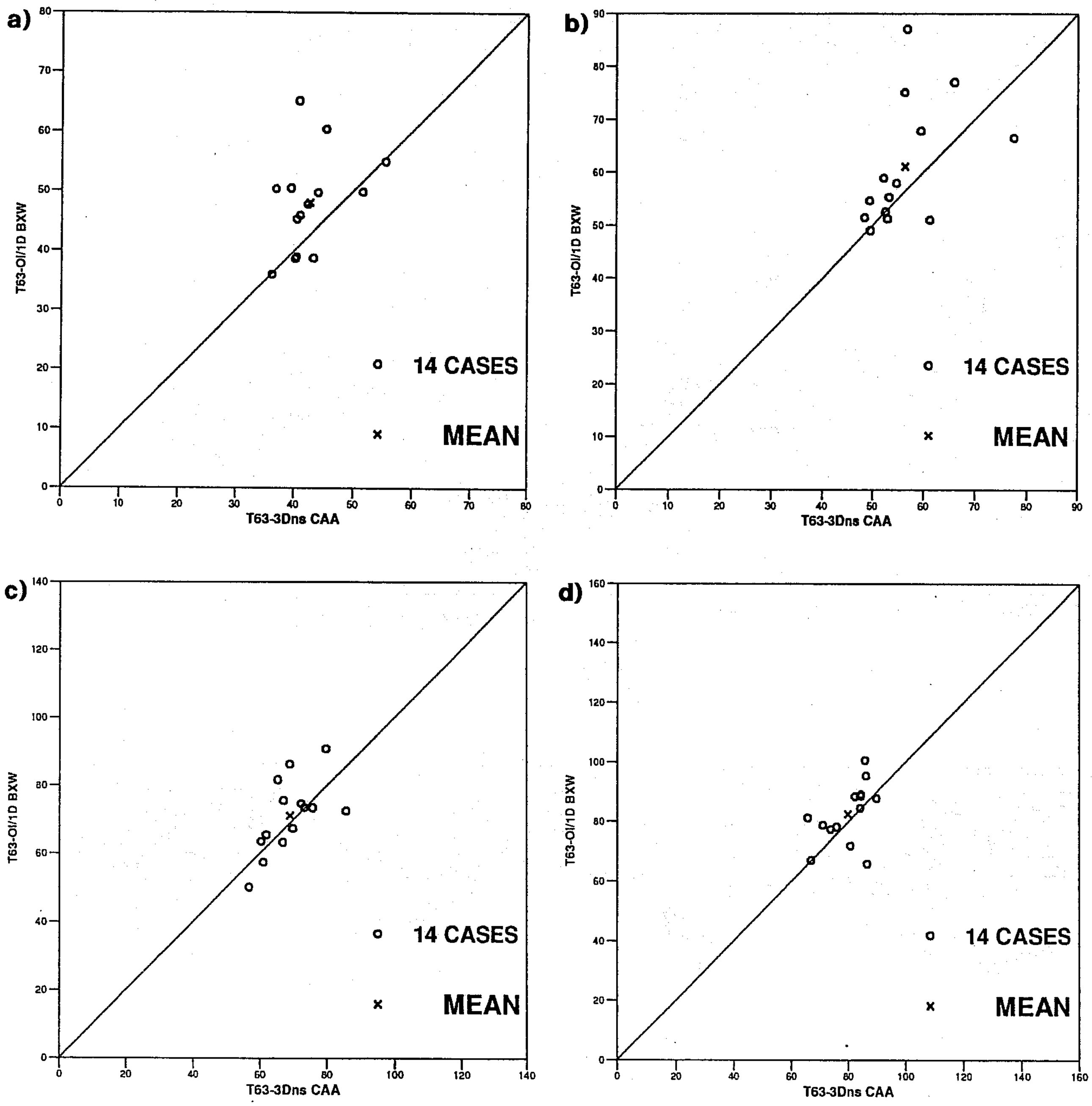

Fig. 27(C) Scatter diagram of the rms of the 500 hPa forecast error issued from 3D-Var (horizontal) and from OI/1D-Var (vertical) verified against the operational 500 hPa height analyses: Southern Hemisphere.
a) 3 day range, b) 4 day range, c) 5 day range, d) 6 day range.

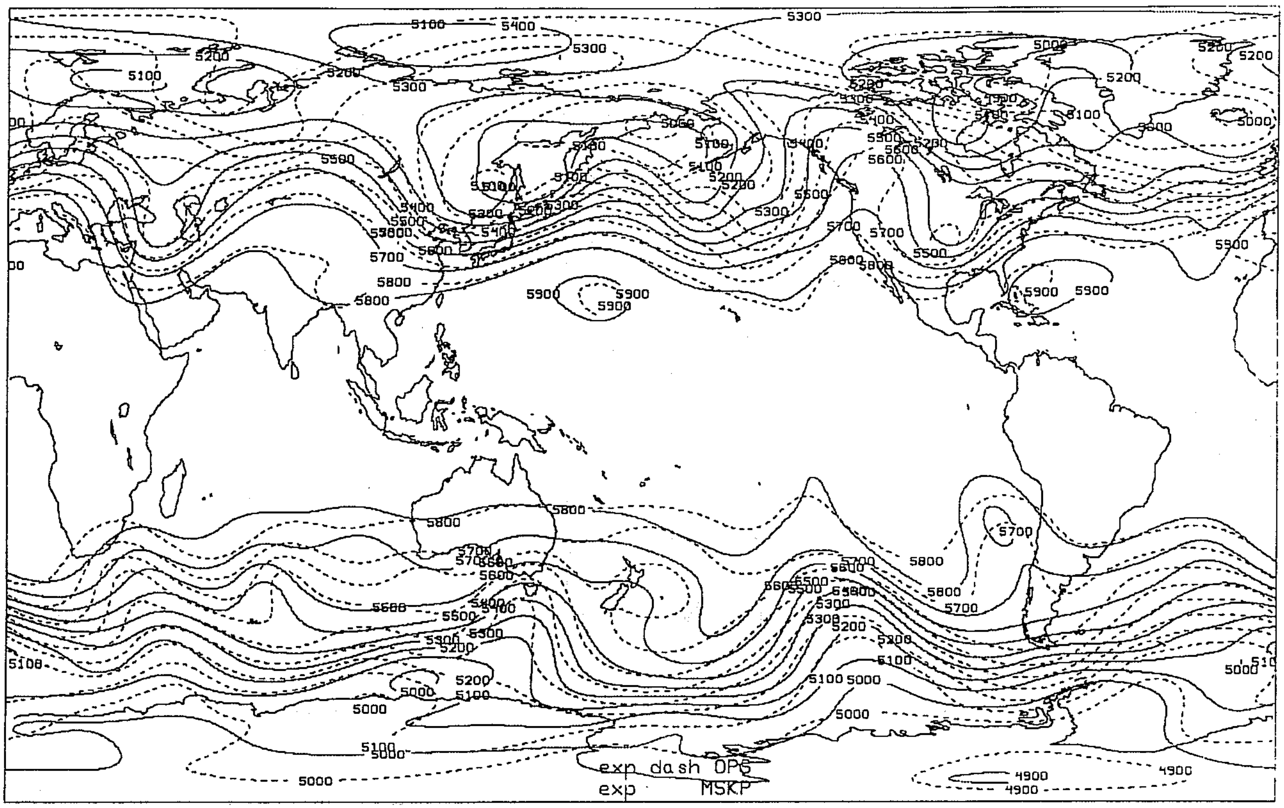

Fig. 28a) 500 hPa geopotential height. Solid: 6 day forecast, dashed: verifying analysis, 3D-Var.

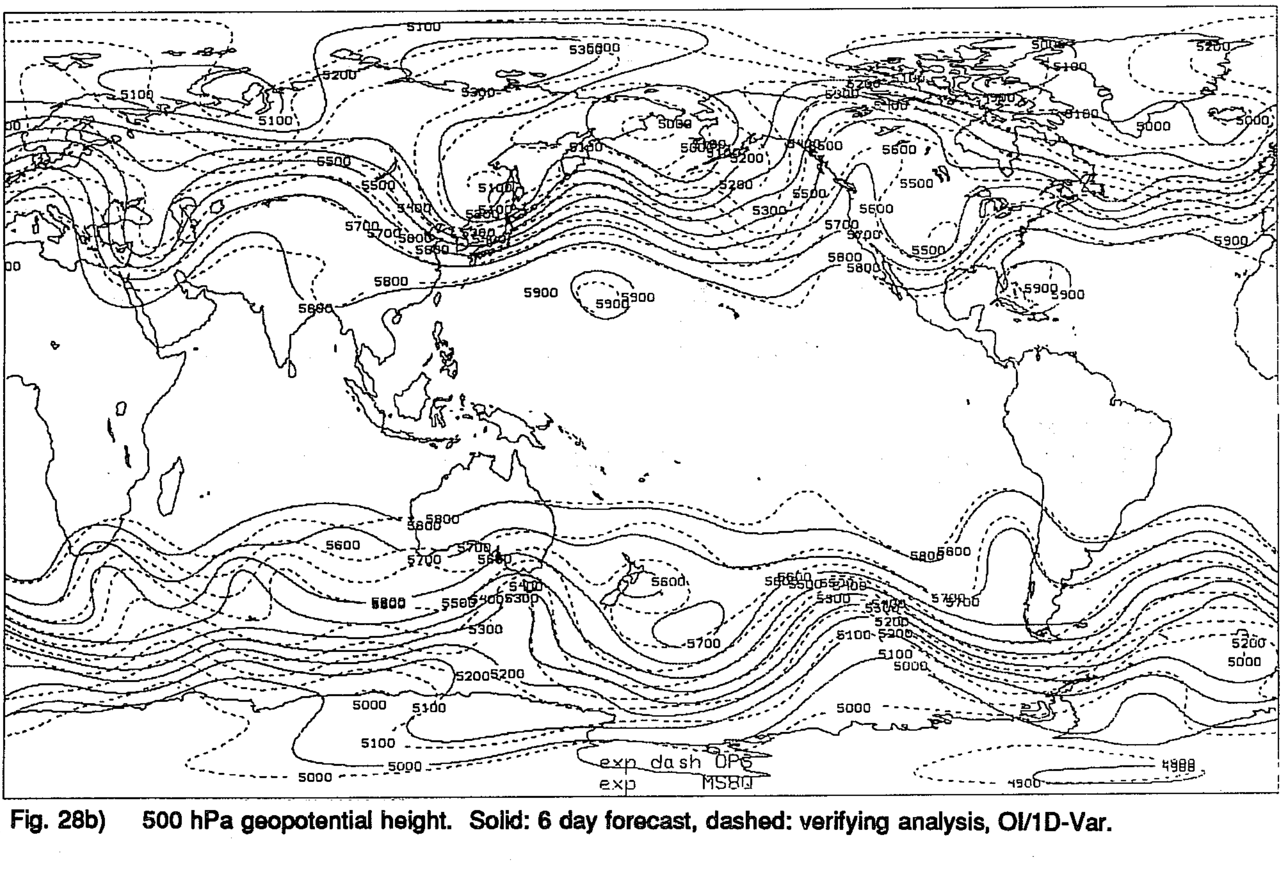

Fig. 28b) 500 hPa geopotential height. Solid: 6 day forecast, dashed: verifying analysis, OI/1D-Var.

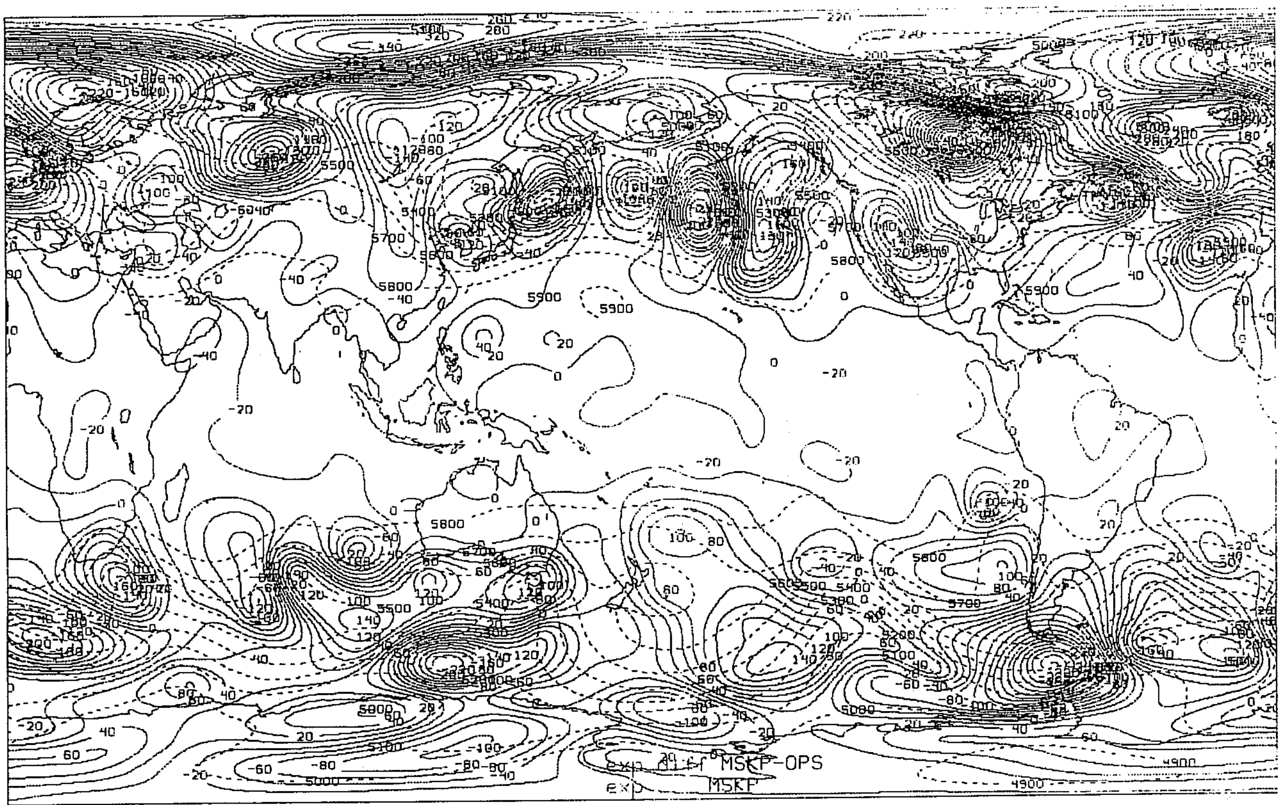

Fig. 29a) 500 hPa geopotential height. 6 day forecast error (solid), dashed: 6 day forecast, 3D-Var.

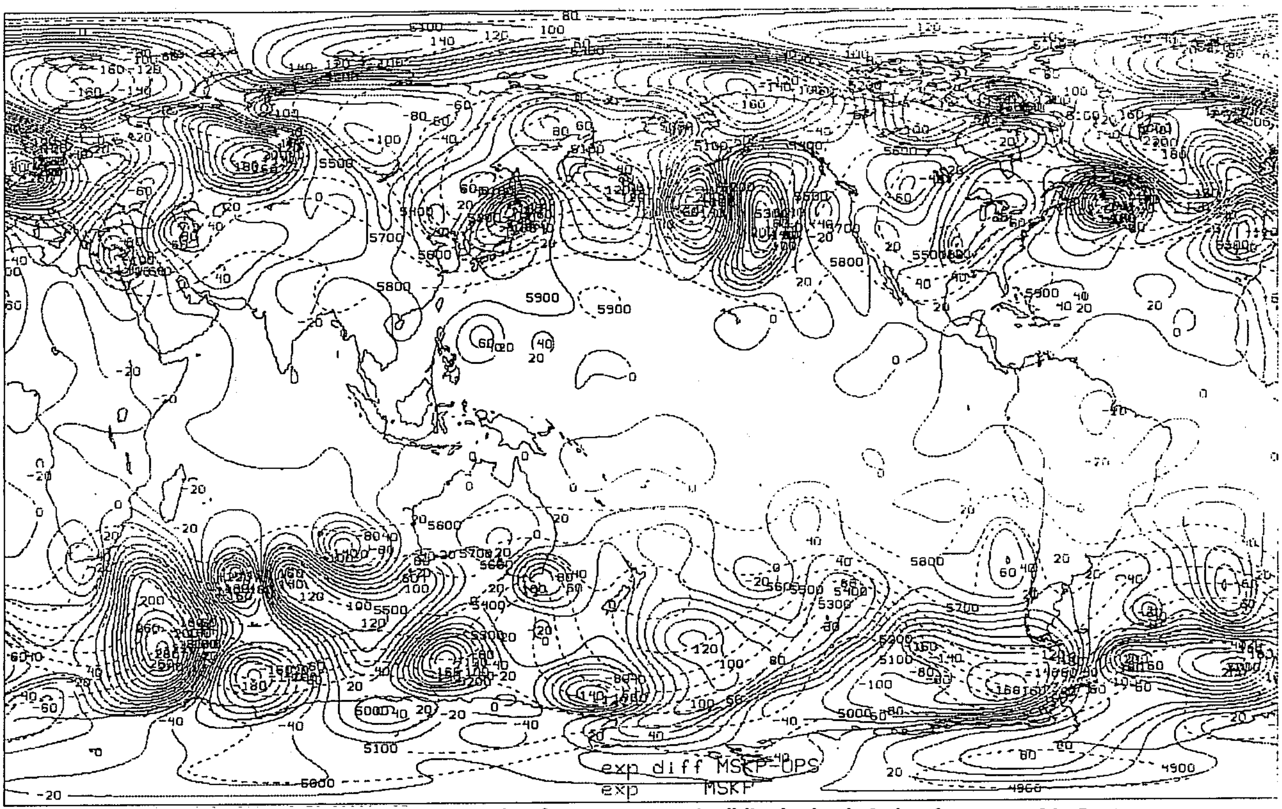

Fig. 29b) 500 hPa geopotential height. 6 day forecast error (solid), dashed: 6 day forecast, OI/1D-Var.

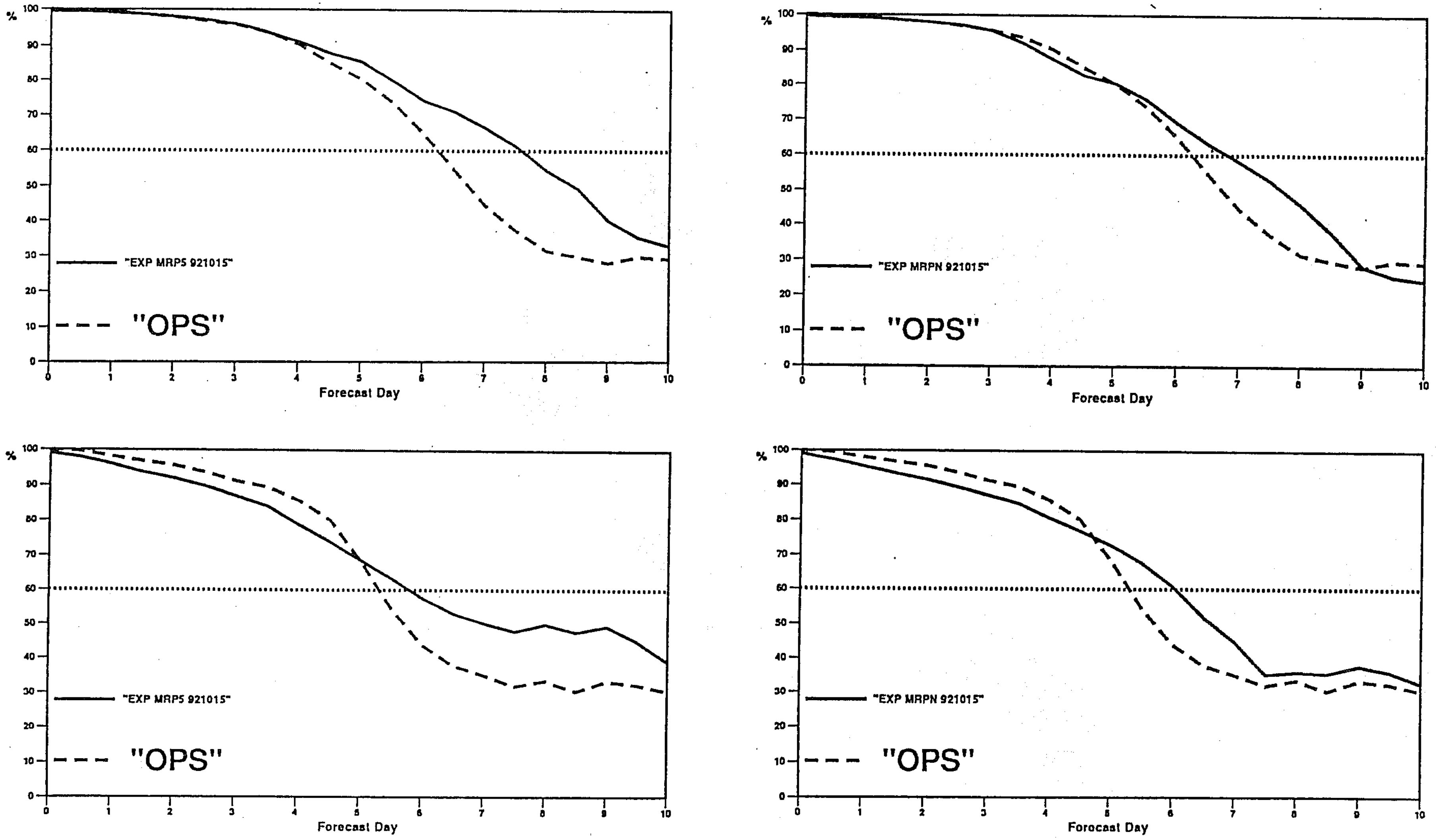

Fig. 30   500 hPa anomaly correlation scores of the forecasts performed from 4D-Var (left) and 3D-Var (right) for Northern Hemisphere (top) and Southern Hemisphere (bottom).

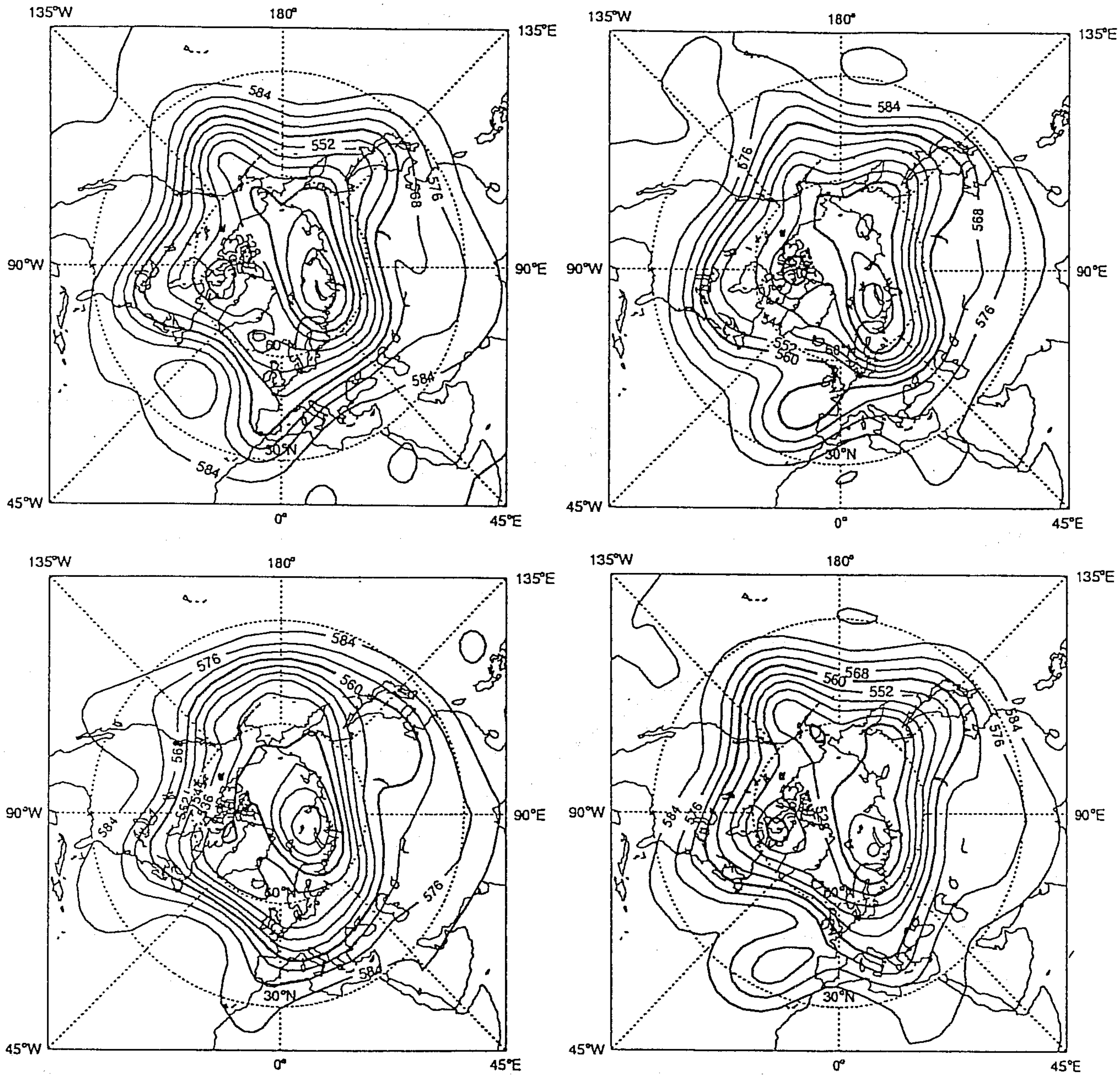

Fig. 31 Average from day 6 to 10 of the 4D-Var forecast (top left), the operational forecast (top right), the 3D-Var forecast (bottom right) and the verifying analysis (bottom left).

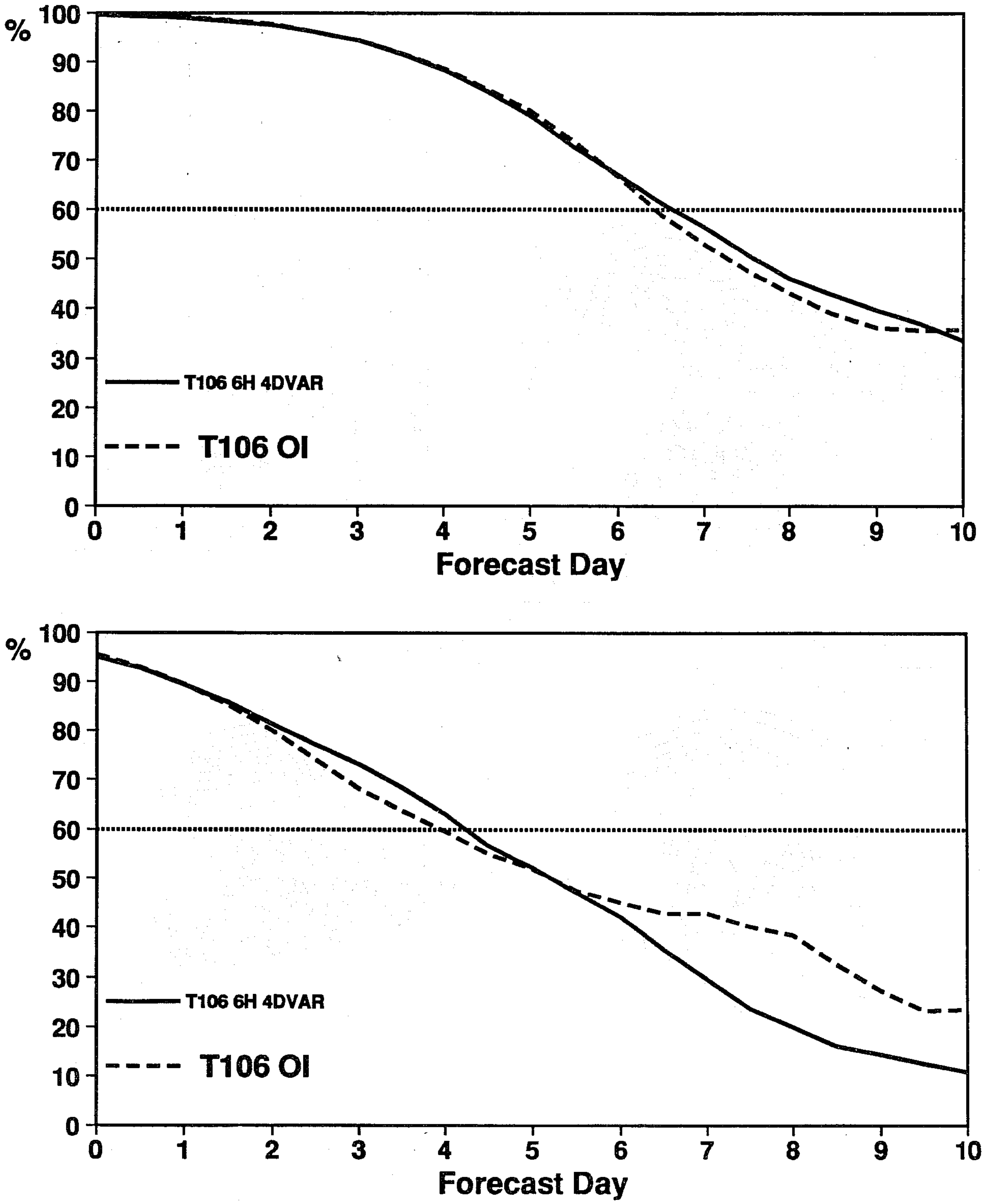

Fig. 32 Average over 6 cases of the forecasts issued from 4D-Var assimilation (solid) and OI (dashed). Top: Northern Hemisphere, bottom: Southern Hemisphere.