TECHNICAL MEMORANDUM

238

The 1997 revision of the Jb term in 3D/4D Var

F. Bouttier, J. Derber and M. Fisher

Research Department

July 1997

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

Abstract

The background term J_b currently used in the 3D-Var and 4D-Var analysis algorithm relies on a generalized linear balance operator. The implied structure functions of the analysis are multivariate ; they depend on the horizontal scale, on the vertical coordinate and on latitude, so that they are very different in the midlatitudes than in the tropical regions. The balance operator is incorporated into the preconditioner of the variational analysis which is now carried out in terms of vorticity, specific humidity, and unbalanced divergence, temperature and surface pressure. The background error covariance model is a non-separable one, as in the previous J_b formulation, and it is expressed in the space of the variables defined by the balance operator. We examine in detail the structure of the J_b operators and their implications for the structure of the analysis increments.

Contents

1 Introduction

The structure of the analysis increments in 3D-Var is driven by the formulation of the so-called background term J_b in the cost-function of the variational analysis. The J_b formulation is very important in 4D-Var as well. An optimal J_b design would ideally reflect the covariances of the short-range forecast errors in the assimilation. In practice, it must be modelized in order to approximate the average variances, autocorrelations and balance properties of the background errors. Between January 1996 and May 1997 the ECMWF 3D-Var operational assimilation system (as well as the experimental 4D-Var) relied on a formulation developed by Heckley et al (1992) and documented in Courtier et al (1997), Andersson et al (1997) and Rabier et al (1997). In May 1997 a new J_b formulation was implemented (corresponding to IFS cycle 16r2), which performs significantly better in 3D-Var and in 4D-Var. The purpose of this paper is to document the revised J_b formulation and to explain its advantages in meteorological, technical and scientific terms.

The new J_b formulation is very different from the previous one, mainly in the tropical structure functions of mass and in the mass/wind coupling of the analysis increments. The specific humidity analysis has not been modified. In the midlatitudes, the new J_b behaves approximately as the previous one, except near the ground and near the tropopause where the geostrophic coupling is now weaker. In the tropical regions, the structure functions of temperature and surface pressure are completely different from the previous ones (particularly the vertical structures), and they have a notable impact on the tropical wind analysis. The autocovariances are non-separable, with a dependence of the vertical covariances on the horizontal total wavenumber n, just as in the previous formulation, but they are not expressed in terms of the same variables, so that although the vorticity autocovariances are assumed to be homogeneous and isotropic, the effective vertical correlations of temperature, for instance, depend on latitude. The multivariate coupling is basically a geostrophic mass/wind linear balance, as in the previous formulation, but it does not rely on the Hough normal modes of the model[1]. Still, it retains all the existing scientific features : there is less geostrophic balance in the smaller horizontal and vertical scales, and virtually no geostrophic balance is assumed near the equator. There are even some new features : the intensity of the geostrophy depends on the vertical level, and there is a weak coupling between divergence and vorticity, as well as between divergence and mass, which may be important in the tropics. Still, there is no tunable parameter in J_b[2], which greatly facilitates its adaptation to new model geometries.

This memorandum is organized in 6 parts :

- a qualitative presentation of J_b, aimed at the non-specialist,
- a mathematical documentation of the formulation,
- a description of its technical implementation,
- an inventory of the information contained in J_b,
- some more elaborated diagnostics of the structure functions,
- a documentation in meteorological terms.

[1]The Hough modes are still used in the J_c term and in the resolution changes needed by the incremental formulation, but not in J_b.

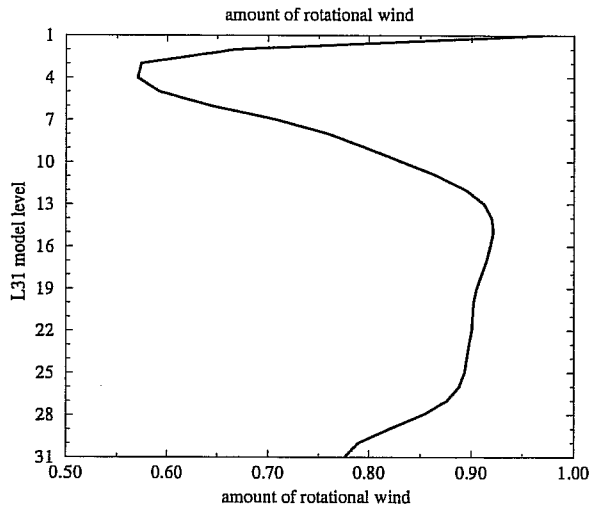
[2]There are still some tunable parameters in the specification of the background standard errors, but not in the spectral covariance model itself.

2 Meteorological features

In this section we explain what the J_b term does in 3D-Var using only basic meteorological concepts. Please refer to the other sections for a rigorous description. The J_b term determines how the 3D-Var analysis procedure converts the differences between observed and forecast variables into corrections of the meteorological fields.

The specific humidity q is analysed independently[3] from the other variables. The structures are globally homogeneous and isotropic, and the analysis weights[4] depend on the first-guess humidity field. There is no vertical correlation above 100hPa.

The wind is analysed in terms of vorticity and divergence, which for the most part are assumed to be independent from each other. The relative amounts of rotational and divergent wind depend on level, they are almost independent of latitude, and they are usually close to 90% rotational. The exact amounts are represented in figure 1.



L31 level	pressure (hPa)	temperature (K)	height (m)
1	10	216	30782
2	30	216	23815
3	50	216	20575
4	70	216	18442
5	90	216	16842
6	110	216	15537
7	132	216	14400
8	155	216	13368
9	181	216	12411
10	208	216	11515
11	238	218	10670
12	270	224	9858
13	304	229	9074
14	339	234	8318
15	377	238	7591
16	416	243	6892
17	457	247	6220
18	500	251	5574
19	543	255	4954
20	588	259	4359
21	633	263	3789
22	679	267	3244
23	725	270	2725
24	772	273	2234
25	817	276	1773
26	861	279	1348
27	902	281	964
28	939	284	628
29	971	285	352
30	995	287	147
31	1009	287	32

Figure 1: Relative amount of rotational wind in the total wind increments in the analysis (as implied by the ratios of background error variance), as a function of L31 model level, and definition of the L31 model levels for a standard atmosphere.

The vorticity structures are globally homogeneous and isotropic. The analysis weights are determined by an assumed tridimensional distribution of background standard error which is the product of a fixed average vertical profile (shown in figure 11) by some horizontal error patterns (an example is displayed in fig.2). The patterns are computed according to a *cycling* algorithm (Fisher and Courtier 1995) that aims to reflect the recent history of the observing network : the background error variance is assumed to be smaller in data-rich areas than in data voids, where it is close to the climatological forecast error. The vorticity analysis is linked to divergence, temperature and surface pressure through

[3]However, there are some links between temperature and specific humidity in the observation operators.

[4]The exact weights depend on the observing network as well, in this section we just describe how they are affected by the definition of J_b.

a complex balance relationship : vorticity corrections imply corrections to the other variables, and vice versa. It implies that the vorticity analysis is important for the analysis of the balanced part of all meteorological variables.

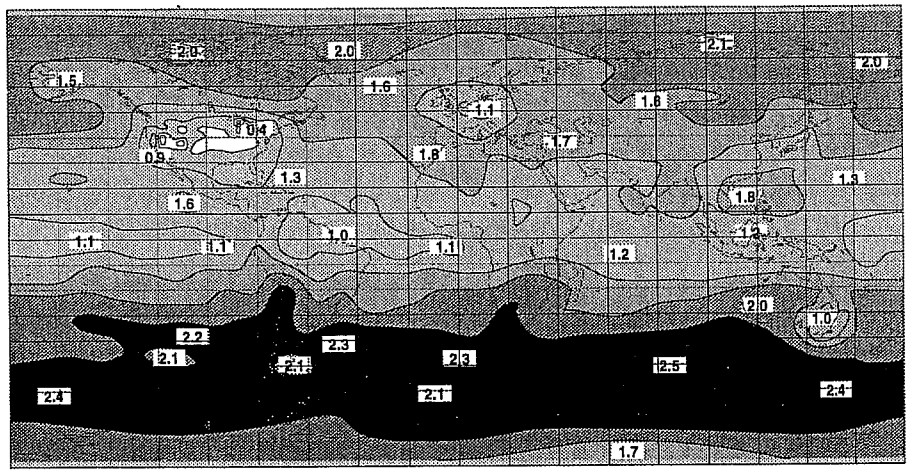


Figure 2: Example of assumed background standard error field distribution for vorticity at level 500hPa on 4th July 1997 at 1800UTC. The unit is 10^{-5}S.I.

The divergence is weakly linked to vorticity. About 90% of the divergence analysis is independent from vorticity, according to "unbalanced divergence" statistics, which are homogeneous and isotropic. The horizontal and vertical correlation structures are sharper for divergence than for vorticity. The assumed background errors of unbalanced divergence are horizontally uniform, and obey a vertical profile which is such that 90% of the wind is rotational at most levels (less in the stratosphere and in the lower troposphere), as shown in figure 1.

The balance of divergence with vorticity accounts for only 10% of the total divergence at most levels, except near the surface and near the tropopause where it is about 30% of the total divergence. It exists only in extratropical regions. The divergence increments are thus the sum of the "unbalanced" divergence analysis everywhere and a small "balanced" divergence analysis implied by the extratropical vorticity increments. The link is such that a cyclonic wind increment generates a weakly convergent wind structure near the ground, and a weakly divergent wind near the tropopause (see fig.31 and 32). The amplitude of the link as a function of level, for a given latitude, is shown in figure 19.

The temperature and surface pressure (i.e. mass) are linked with the wind. The nature of the coupling is very different in the tropical and in the extratropical regions, and this has a large impact on the (T, p_s) structures and analysis weights.

The extratropical (T, p_s) analysis is almost completely balanced with vorticity in the extratropical regions : there, (T, p_s) increments will be accompanied by strong vorticity increments and vice versa. In the midlatitudes (around 50N and 50S), 80% of T and p_s are geostrophically linked to vorticity ; the geostrophic coupling on temperature is weaker (down to about 50%) near the surface and near the tropopause. There is a very weak link between (T, p_s) and divergence as well. The remaining 20% of T and p_s are analysed independently from the other variables. This implies that the analysis weights of T and P_s are predominantly determined by the assumed standard errors for vorticity. The

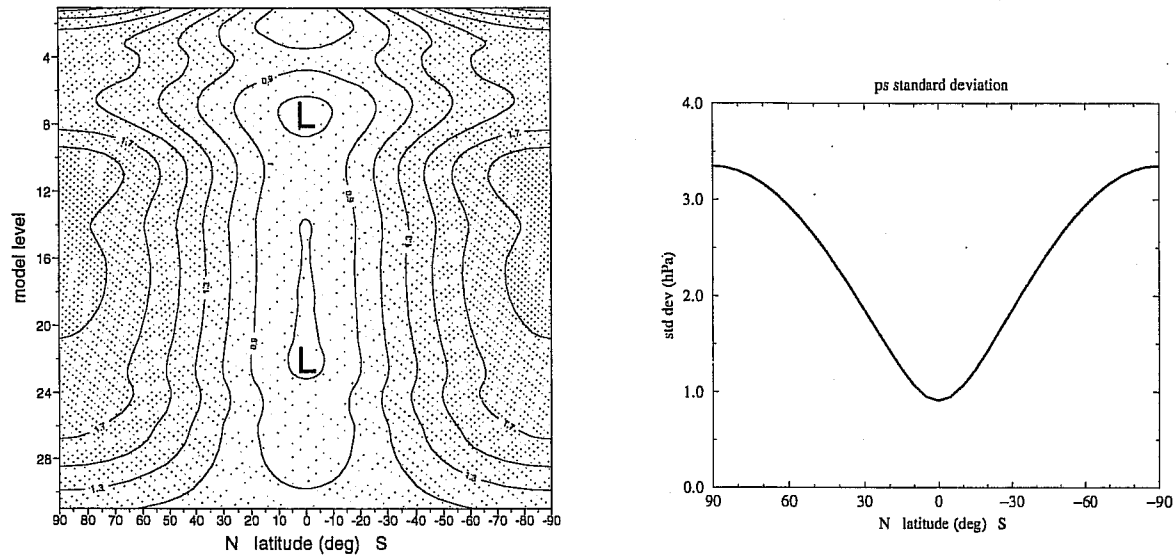


Figure 3: Latitudinal dependence of background errors of temperature (left panel) for all model levels, and of surface pressure (right panel), before the horizontal modulation of the vorticity errors by the cycling algorithm patterns.

geostrophic coupling is weaker for the small horizontal and vertical features of the fields than for the large scales. The (T, p_s) increments are the sum of the vorticity-balanced, divergence-balanced and unbalanced (T, p_s) increments ; their relative magnitude as a function of height is shown in figure 19.

The tropical (T, p_s) analysis has almost no geostrophic coupling with vorticity (there is none at the equator). There is a coupling between (T, p_s) and divergence such that 30% of T is coupled near the ground and near the tropopause ; at the other levels it is only 10%, as shown in figure 19. An increase in temperature (or a decrease in p_s) is associated with weak convergence near the ground, and weak divergence near the tropopause (see fig.30 for an example). Hence, most of the (T, p_s) analysis is independent from the wind, and the vertical temperature structures are much sharper than in the midlatitudes ; they are almost identical to those implied by the "unbalanced" (T, p_s) covariances.

the analysis weights for (T, p_s) are completely different near the equator and in the extratropics (unlike wind). As explained above, most of the extratropical weights are implied by the vorticity weights through the geostrophic balance. The variation of the implied background errors with latitude is displayed in figure 3 ; it is the sum of a uniform unbalanced part, and a balanced part which is modulated by the horizontal patterns of vorticity error ; the cycling implies larger errors in the Southern than in the Northern Hemisphere, and local error maxima around midlatitude jets.

the vertical correlations between T **and** p_s are generally negative ; they are about -0.3 near the ground, and very small in the mid-troposphere. In the extratropics, there is also a -0.25 negative correlation around the tropopause.

Some more quantitative information will be given in the following sections, with the exact correlation structures and standard deviation vertical profiles, as well as some example of analysis increments with simulated observations.

3 Description of the algorithm

We use the following notations :

δx : low-resolution analysis increment, i.e. model field departures from the background,

B : assumed background error covariance matrix,

ζ, η, (T, p_s), q : respectively, increments of vorticity, divergence, temperature and surface pressure, and specific humidity, on model levels.

η_b, $(T, p_s)_b$: **balanced** parts of the η and (T, p_s) increments. The concept of balance will be defined below.

η_u, $(T, p_s)_u$: **unbalanced** parts of η and (T, p_s), i.e. $(\eta - \eta_b)$ and $[(T, p_s) - (T, p_s)_b]$, respectively.

The incremental variational analysis problem

$$\begin{aligned} \min J(\delta \mathrm{x}) &= J_b(\delta \mathrm{x}) + J_o(\delta \mathrm{x}) \\ &= \delta \mathrm{x}^{\mathrm{T}} \mathbf{B}^{-1} \delta \mathrm{x} + J_o(\delta \mathrm{x}) \end{aligned}$$

is rewritten in the space defined by the change of variable $\delta \mathrm{x} = L\chi$ where L verifies $LL^{\mathrm{T}} = \mathbf{B}$ so that the same analysis increments are provided by

$$\begin{aligned} \min J(\chi) &= \chi^{\mathrm{T}} \chi + J_o(L\chi) \\ \delta \mathrm{x} &= L\chi \,. \end{aligned}$$

In the operational practice, the initial point of the minimization is the background, so that initially $\delta \mathrm{x} = \chi = 0$. The minimization can be carried out in the space of χ, where J_b is the euclidean inner product. At the end of the minimization, the analysis increments are reconstructed in model space by $\delta \mathrm{x} = L\chi$. Thus the variational analysis can be done with L, the change of variable from minimization space to model space (`chavarin` in the IFS/Arpège code), without ever using its inverse.

The background error covariance model **B** is implied by the design of L, which in the current J_b formulation has the form

$$L = K \mathbf{B}_u^{1/2}$$

where K is a **balance** operator going from the set of variables $[\zeta, \eta_u, (T, p_s)_u, q]$ to the model variables $[\zeta, \eta, (T, p_s), q]$. The $\mathbf{B}_u^{1/2}$ operator is the right-hand symmetric square root of the background error covariances $\mathbf{B}_u$ of $[\zeta, \eta_u, (T, p_s)_u, q]$, so that

$$\mathbf{B}_u = \mathbf{B}_u^{\mathrm{T}/2} \, \mathbf{B}_u^{1/2} \,.$$

So far the formulation is perfectly general. Now we are going to restrict $\mathbf{B}_u$ to a simple form and to choose a particular balance operator K.

The covariance matrix $\mathbf{B}_u$ is assumed to be block-diagonal, with no correlation between the parameters :

$$\mathbf{B}_u = \begin{bmatrix} C_\zeta & 0 & 0 & 0 \\ 0 & C_{\eta_u} & 0 & 0 \\ 0 & 0 & C_{(T,p_s)_u} & 0 \\ 0 & 0 & 0 & C_q \end{bmatrix}$$

It implies that the q analysis is independent from the other variables. However, assuming that the unbalanced variables are uncorrelated is not too restrictive because, as we shall see below, the design of the balance implies significant multivariate correlations between the meteorological variables.

Each autocovariance block in the above matrix is itself assumed to be block-diagonal in spectral space, with no correlation between different spectral coefficients, but a full vertical autocovariance matrix for each spectral coefficient. The vertical covariance matrices are assumed to depend only on the total wavenumber n. The resulting autocovariance model is homogeneous, isotropic and non-separable in gridpoint space : the correlation structures do not depend on the geographical location, but they depend on the scale. The shape of the horizontal correlations is determined by the covariance spectra. The same representation was used in the previous J_b formulation (Rabier and McNally 1993, Courtier et al 1997). The covariance coefficients are computed statistically using the NMC method (Parrish and Derber 1992, Rabier et al 1997) on 24/48-hour forecast differences to estimate the total covariances for each total wavenumber n, and assuming an equipartition of errors between the $(2n+1)$ associated spectral coefficients.

The balance relationship is arbitrarily restricted to the following form :

$$\begin{aligned} \eta_b &= M\zeta \\ (T,p_s)_b &= N\zeta + P\eta_u \end{aligned}$$

So that the complete balance operator K is defined by :

$$K : [\zeta, \eta_u, (T,p_s)_u, q] \mapsto [\zeta, \eta, (T,p_s), q]$$

$$\begin{aligned} \zeta &= \zeta \\ \eta &= M\zeta + \eta_u \\ (T,ps) &= N\zeta + P\eta_u + (T,ps)_u \\ q &= q \end{aligned}$$

or equivalently, in matrix form :

$$K = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ M & \mathbf{I} & 0 & 0 \\ N & P & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix}$$

The matrix blocks M, N, P are in general not invertible, but K is :

$$K^{-1} = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ -M & \mathbf{I} & 0 & 0 \\ (PM-N) & -P & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix}$$

As explained above, the inverse of K is not actually used in the variational analysis, because the initial point of the minimization is the background.

The matrix multiplication of $\mathbf{B}_u$ by K allows one to write explicitly the implied background error covariance matrix $\mathbf{B}$ in terms of the meteorological variables $[\zeta, \eta, (T, p_s), q]$:

$$\mathbf{B} = K\mathbf{B}_u K^{\mathrm{T}} = \begin{bmatrix} C_\zeta & C_\zeta M^{\mathrm{T}} & C_\zeta N^{\mathrm{T}} & 0 \\ MC_\zeta & MC_\zeta M^{\mathrm{T}} + C_{\eta_u} & MC_\zeta N^{\mathrm{T}} + C_{\eta_u} P^{\mathrm{T}} & 0 \\ NC_\zeta & NC_\zeta M^{\mathrm{T}} + PC_{\eta_u} & NC_\zeta N^{\mathrm{T}} + PC_{\eta_u} P^{\mathrm{T}} + C_{(T,p_s)_u} & 0 \\ 0 & 0 & 0 & C_q \end{bmatrix}$$

The blocks implied by C_ζ and its transforms by the balance operator blocks M, N, P are the "balanced" parts of the covariances. For instance, the vorticity covariances C_ζ and the unbalanced temperature covariances $C_{(T,p_s)_u}$ are both homogeneous and isotropic, whereas the $NC_\zeta N^{\mathrm{T}}$ "vorticity-balanced" (T, p_s) matrix term depends on latitude : it is predominant in the extratropics, negligible near the equator. The NC_ζ term is responsible for the geostrophic mass/wind coupling.

The M, N and P operators used to define the balance have a restricted algebraic structure. M and N are both the product of a so-called horizontal balance operator $\mathcal{H}$ by vertical balance operators $(\mathcal{M}, \mathcal{N})$:

$$\begin{aligned} M &= \mathcal{M}\mathcal{H} \\ N &= \mathcal{N}\mathcal{H} \end{aligned}$$

The $\mathcal{H}$ operator is a block-diagonal matrix of identical horizontal operators transforming the spectral coefficients of vorticity, independently at each level, into an intermediate variable P_b which is a kind of linearized mass variable defined below. The horizontal operators in $\mathcal{H}$ have exactly the same algebraic structure as the standard analytical linear balance on the sphere, and this is where the latitudinal variations of the J_b structures come from : in spectral space,

$$P_b(n,m) = \beta_1(n,m)\,\zeta(n,m+1) \;+\; \beta_2(n,m)\,\zeta(n,m-1)$$

The $\mathcal{M}$, $\mathcal{N}$ and P operators all have the same structure : block-diagonal, with one full vertical matrix per spectral component. The vertical matrices depend only on the total wavenumber n.

The actual calibration of the J_b operator requires the following 4 steps ; each one uses a set of 24/48h-range forecast differences as surrogates to background error patterns in order to calculate the statistics :

step 1 : $\mathcal{H}$ operator. The horizontal balance coefficients (β_1, β_2) of $\mathcal{H}$ are computed by a linear regression between the errors in vorticity and in linearized total mass P_t, assuming the functional relationship defined by the above equation, and building P_t from (T, p_s) using the linearized hydrostatic relationship : at level l,

$$P_t(l) = \sum_{i=\mathrm{NFLEV}}^{l} RT_i\,\Delta \log p_i + RT_{\mathrm{ref}} \log p_s$$

which relies on the definition of the model vertical geometry and of "reference" values for (T, p_s) ; we use (270K, 800hPa) currently[5].

[5]The sensitivity to the somewhat arbitrary choice of these parameters has been tested and it is negligible. Unlike in the previous J_b formulation, P_b is just an intermediate variable in the linear regression. Modifying the reference values, e.g. to (300K,1000hPa), does change the scaling of $\mathcal{H}$, but it is compensated by corresponding changes in the $\mathcal{M}$ and $\mathcal{N}$ operators, so that the effective covariances are virtually unchanged.

step 2 : $\mathcal{M}$ operator. The vertical blocks $\mathcal{M}(n)$ of this operator are computed for each wavenumber n by a linear regression between the spectral vertical profiles $[P_b]_n^m$ and $[\eta]_n^m$, respectively, of balanced mass P_b (defined as $\mathcal{H}$ times the vorticity error patterns) and divergence. The relationship is assumed to be

$$[\eta]_n^m = \mathcal{M}(n)[P_b]_n^m$$

so that the statistical sampling is better for the small scales than for the large scales because there are $(2n+1)$ spectral profiles to be used per total wavenumber in each forecast error pattern. At least as many independent error patterns as number of model levels are needed in order to have a well-posed regression problem for the very large scales.

step 3 : $\mathcal{N}$ and P operators. The vertical blocks are computed for each wavenumber exactly like $\mathcal{M}$, except that now the linear regression goes from the vertical spectral profiles of $P_b = \mathcal{H}\zeta$ and $\eta_u = \eta - M\zeta$ to the profiles of temperature concatenated with surface pressure:

$$[(T, p_s)]_n^m = \mathcal{N}_n[P_b]_n^m + \mathcal{P}_n[\eta_u]_n^m$$

One notes that the $\mathcal{N}_n$ matrix is not square (the output is larger than the input because there is a kernel in the hydrostatic relationship) but the resulting (T, p_s) covariances are still positive definite by construction thanks to the $C_{(T,p_s)_u}$ term in the expression of $\mathbf{B}$.

step 4 : Error covariances. The vertical autocovariances of the $[\zeta, \eta_u, (T, p_s)_u, q]$ difference patterns are computed for each total wavenumber n. Again, since there are $(2n+1)$ wavenumbers for each n and each error pattern, at least as many linearly independent error patterns as model levels (plus one for p_s) *must* be used in order to ensure that the autocovariances are positive definite at the very large scales. It is strongly advised to use several times more in order to reduce the sampling noise at large scales ; this is important for the performance of the resulting assimilation/forecast systen. In the May 1997 implementation of the 3D-Var system, about 180 forecast difference patterns have been used for 31 levels.

In addition to these 4 steps, some minor preprocessing is performed on the covariances. The vertical correlations of humidity are set to zero above 100hPa in order to avoid spurious stratospheric humidity increments because of the tropospheric observations. The ζ, η_u and $(T, p_s)_u$ vertical profiles of total variance are rescaled by an arbitrary factor of 0.9 in order to account for the mismatch between the amplitudes of the 24/48-hour forecast differences and of the 6-hour forecast errors. In the future this factor will be recalculated more precisely using observation departures from the background in the assimilation, similarly to Hollingsworth and Lönnberg (1986). It may be different for 3D-Var than for 4D-Var. The variance spectra are slighty modified in order to ensure that the horizontal error correlations of ζ, η_u and (T, p_s) are compactly supported (they are set to zero beyond 6000km) ; this operation removes the residual sampling noise in the error covariances. No other processing is performed except for a spectral truncation if the analysis resolution is lower than the statistics resolution (currently T106). It would be easy to extrapolate the statistics to higher resolutions, but it would be very hazardous to alter the vertical geometry of the covariances and balance operators. Instead, it is recommended to run a set of forecasts using a model with the right vertical resolution, and recompute all the statistics from scratch.

4 Technical implementation

The statistical calibration is done using dedicated scripts outside the IFS/Arpège code. First, the 24/48-hour forecast error differences for a set of dates are constructed in terms of spectral (ζ, η, T, p_s, q). At ECMWF this involves running a set of MARS requests and building the required `GRIB` files. Then, the forecast error differences are read and processed by a `Fortran` statistics program that finally writes two files in `GSA` format : one with the coefficients of the balance operator, one with the error covariances of $[\zeta, \eta_u, (T, p_s)_u, q]$. These files take up a couple of megabytes. They are computed for a given triangular truncation `nsmax` and number of levels `nflev` ; currently `nsmax`=106 and `nflev`=31 . In the covariance file there are 4 sets of `nsmax` vertical covariance matrices, each of dimension `nflev`×`nflev` except the $(T, p_s)_u$ ones which are at `(nflev+1)`×`(nflev+1)`. The balance files contain one set of twice `(nsmax+1)`×`(nsmax+2)` coefficients for the $\mathcal{H}$ operator, and three sets of `nsmax` vertical balance matrices of dimension `nflev`×`nflev` for $\mathcal{M}$, `(nflev+1)`×`nflev` $\mathcal{N}$ and P.

The IFS needs these two `GSA` files to use the J_b code (e.g. in the incremental analysis jobs). The J_b configuration described here corresponds to namelist switch `LSTABAL=.TRUE.`, and it is identified in the J_b code by the string `CDJBTYPE='STABAL96'`. The input files must be named `stabal96.bal` and `stabal96.cv` . Some important namelist options are `LCFCE` (to enforce uniform background errors), `L3DBGERR` (to have a 3-D distribution vorticity background errors), and `LCORCOSU` (to enforce compactly supported horizontal correlations). Many options of the previous J_b (including stretching) are obsolete or not currently supported.

Inside the IFS code, J_b is localized in the setups below subroutine `subjcov` and in the inverse change of variable `cvar2in` (and its adjoint and their inverses). The computation of the cost function and its adjoint is done in `sim4d` ; it is planned to move it to a dedicated subroutine. The distributed memory affects the setups below `sujbdat` and `sujbbal` when the data files are read (by the master processor only) : first, the resolution of the files is read, then the relevant arrays are allocated and the actual data is read, truncated if necessary, and broadcast. The code is designed to work at any resolution. In the change of variable, there is a transposition of the fields between the horizontal and vertical balance operators, respectively `balstat` and `balvert`. Note that the operator K is performed by calling `cvar2in`, so in IFS/Arpège parlance K corresponds to the previous *inverse* change of variable.

The background standard errors are set up below `sujbstd` and used in `jgnr` or `jgnrs` (and their adjoint and inverses). On top of the covariance files, they use a gridpoint `GRIB` file called `errgrib` in order to specify the three-dimensional error patterns. The data from the file is converted to the right parameters and resolution if needed. The background error fields for some parameters (wind, height, temperature and surface pressure) are built for the screening job although they are not needed in the analysis itself. For more information, refer to the documentation about the cycling.

The J_b code is substantially simpler and easier to maintain than in the previous formulation. It does not involve the Hough normal modes nor any arbitrary tuning parameter, except a single number to scale the variances. It is computationally cheaper (there are 4 times fewer spectral transforms than before) and it provides a better preconditioning of the minimization because the J_b cost function is spherical in control variable space. Perhaps more importantly, the knowledge of the exact symmetric square root of the background error covariance operator $\mathbf{B}$ and its inverse is going to facilitate a lot the developments of simplified Kalman filters.

5 Information contained in J_b

In this paper, the autocorrelations are always plotted using a contour interval of 0.1, with one contour at the value 0.05 . Unless mentioned in the figure captions, S.I. units are always used. We present statistics triangularly truncated at T106, although the current ECMWF analysis is carried out at T63 only. The L31 vertical model geometry is documented in figure 1 ; it is important to remember that level 18 is at 500hPa, and the tropopause is (very roughly) around model 10, which is at 200hPa.

5.1 Balance operators

Horizontal balance $\mathcal{H}$

This operator is defined by the spectral distribution of coefficients β_1 and β_2 in the relationship

$$P_b(n,m) = \beta_1(n,m)\,\zeta(n,m+1) \;+\; \beta_2(n,m)\,\zeta(n,m-1)$$

where the coefficients outside the truncation domain are set to zero. The above equation has the same structure as the one implied by the geostrophic equilibrium, i.e. the balance between the Coriolis force and the pressure gradient : on pressure surfaces,

$$f.\,k \wedge u = -\nabla\Phi_b$$

or, equivalently, on the sphere :

$$\Phi_b = \Delta^{-1}\mathrm{div}(f.\nabla\Delta^{-1}\zeta)$$

where Φ is the geopotential and f is the Coriolis factor ; as in J_b, we arbitrarily choose to write this constraint as an application from vorticity to Φ_b, the geostrophically balanced part of Φ. In J_b, the balance is applied on model levels, which are close to pressure levels except near the ground and over high orography. The discretization of the above equation leads to the analytical balance operator :

$$\begin{aligned}
\Phi_b(n,m) &= b_1(n,m)\,\zeta(n,m+1) \;+\; b_2(n,m)\,\zeta(n,m-1) \\
b_1(n,m) &= b\frac{1}{n(n+1)}\left(1-\frac{1}{n+1}\right)\sqrt{\frac{(n+1+m)(n+1-m)}{(2n+3)(2n+1)}} \\
b_2(n,m) &= b\frac{1}{n(n+1)}\left(1+\frac{1}{n}\right)\sqrt{\frac{(n+m)(n-m)}{(2n+1)(2n-1)}}
\end{aligned}$$

where $b = -2\Omega R_a^2$, Ω is the rotation speed of the Earth, R_a is its radius. Qualitatively this is the inverse Laplacian $-R_a/n(n+1)$ of vorticity times a corrective term that accounts for β-effects.

We are going to show that the $\mathcal{H}$ operator is almost identical to the analytical balance operator. A first indication of this is in the coefficients themselves. Figure 4 shows the spectral distribution in (n,m) space of the quantities $|\log_{10}(\beta_1/b_1)|$ and $|\log_{10}(\beta_2/b_2)|$ which are measures of the distance between (β_1,β_2) and (b_1,b_2). There is a significant sampling noise, particularly

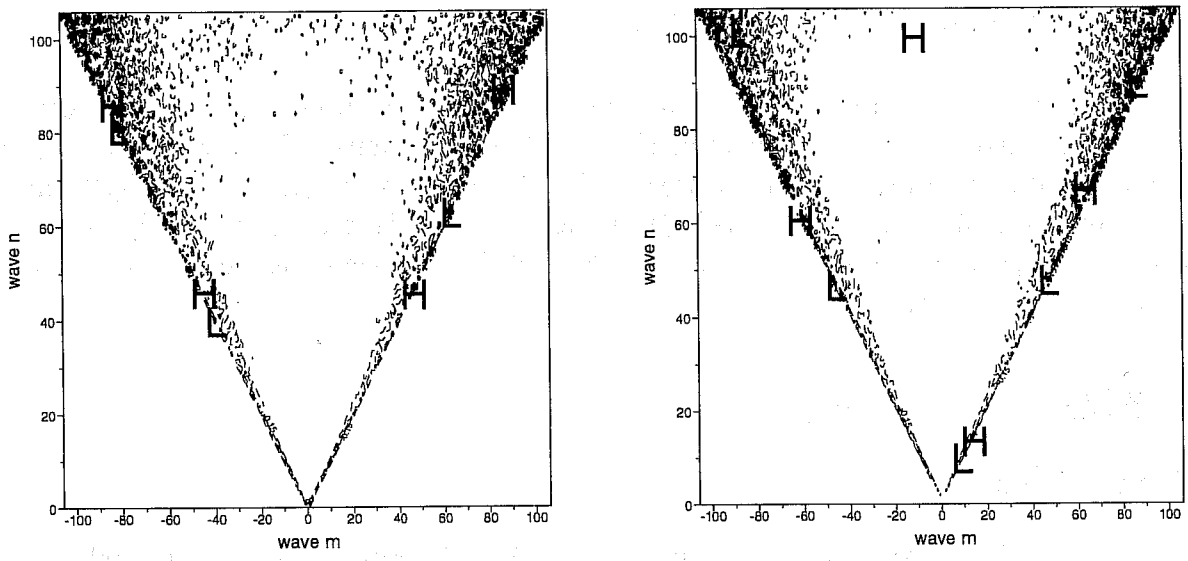


Figure 4: Representation of the distance between the spectral horizontal balance factors $\beta_1(n,m)$ (left panel) and $\beta_2(n,m)$ (right panel), and the corresponding linear balance analytical coefficients, as explained in the text. The ordinate is coefficient n, the abcissa is m (imaginary part on the left, real part on the right). The isoline contour interval is 0.1, i.e. the lowest isoline delineates 5% differences.

in the smaller scales. The larger values at the edges (up to 50% difference) are not understood but they are probably an artifact of the linear regression being carried out in terms of one predictor, instead of two in the rest of the spectral triangle. However there does not seem to be any significant departure of $\mathcal{H}$ from the analytical balance : most of the differences are less than 10%.

Another characterization is provided in figure 5 where the operator $\mathcal{H}$ and the analytical balance have been applied to the same vorticity difference field. The chosen level is 240hPa (model level 11), it is the one at which the largest geopotential height differences have been found for this particular forecast difference field (the day has been chosen at random). The height patterns are almost identical everywhere. The quantitative differences are usually small, less than 1m (the maximum is 1.33m). The height differences produced by $\mathcal{H}$ tend to be smaller, probably because the linear regression is done once for all levels, so that it includes the lower ones for which the geostrophic coupling is weaker than in the free atmosphere.

One might still argue that there is a difference in the equatorial mass/wind coupling. Daley (1997) argued that the use of the analytical balance or Hough modes to enforce geostrophism in variational analysis increments is inappropriate in the tropical regions, and he designed a modified linear balance operator with its pseudo-inverse in order to alleviate the problem (by reducing the mass/wind coupling in the tropical regions). We shall demonstrate below that the problem does not exist in this formulation, but it is not because of $\mathcal{H}$. There would be a problem if the balance constraint in J_b were formulated so that the minimization attempts to invert the balance operator, in order to generate wind increments from equatorial geopotential observation. Actually, both $\mathcal{H}$ and the analytical balance produce very small geopotential increments in the tropical regions, as shown in fig.5, and they are both ill-suited for algebraic inversion. This problem does not exist in the J_b formulation because the presence of the unbalanced $(T, p_s)_u$ autocovariance term makes the analysis univariate near the equator, so that the precise definition of $\mathcal{H}$ has absolutely no importance in the wind or mass analysis there.

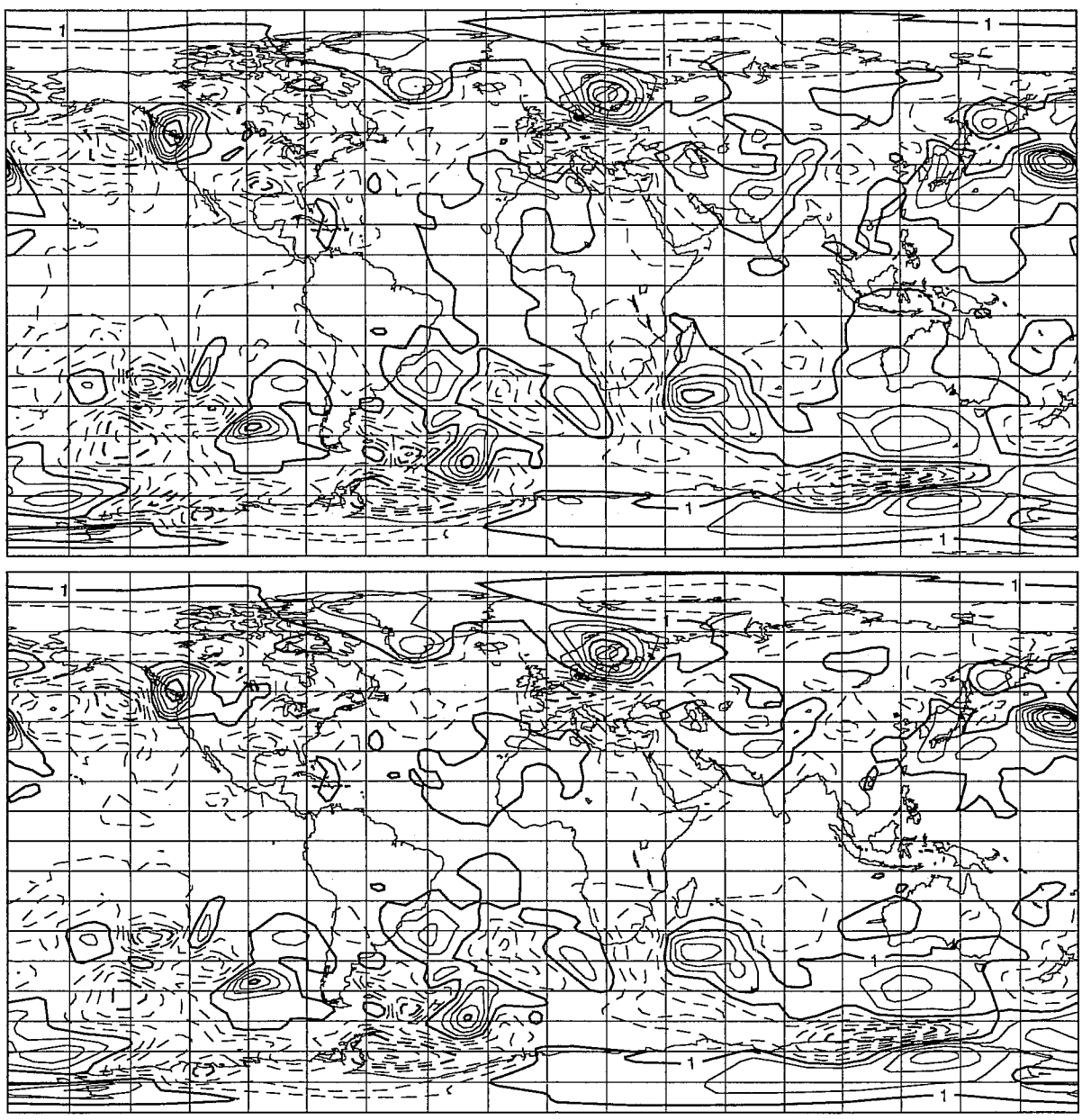


Figure 5: Result of the application of the horizontal balance operators on a vorticity difference field, in terms of balanced geopotential height difference (the isoline contour interval is 1m, with an isoline at 0.5m, and negative isolines are dashed). Top panel : balanced height according to the analytical balance, bottom panel : according to the statistical balance $\mathcal{H}$.

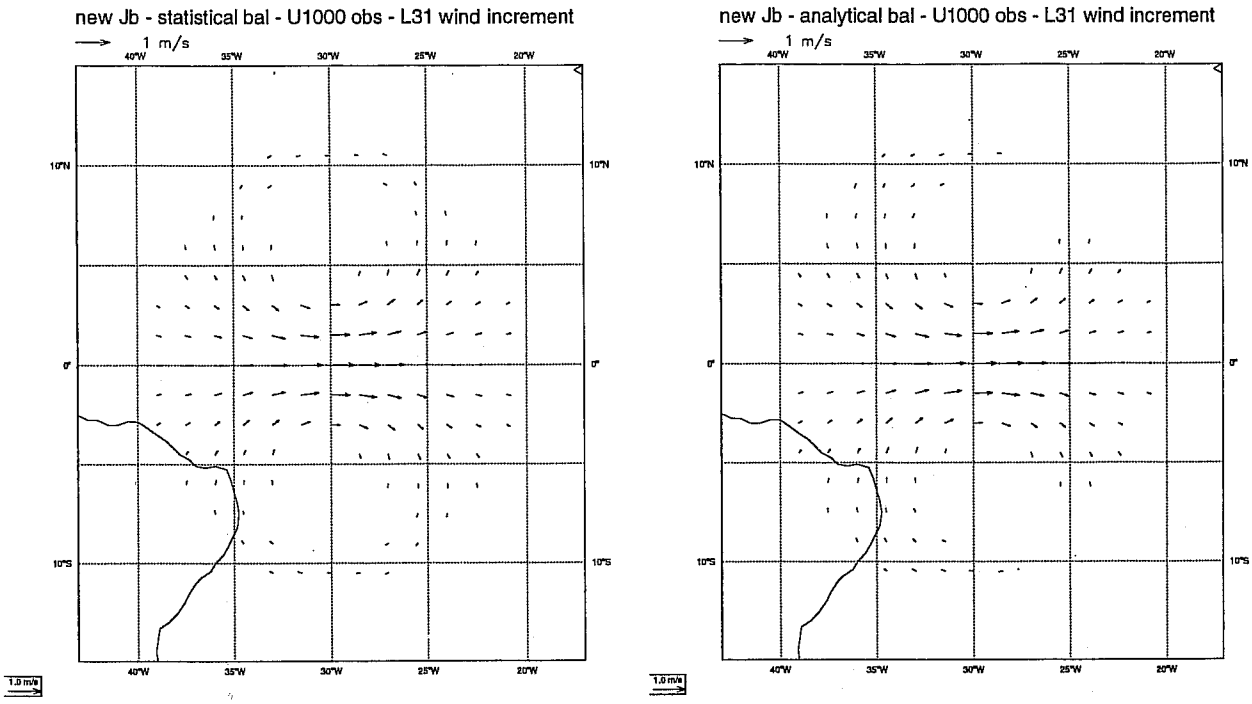


Figure 6: Wind increments at the lowest model level produced by a simulated wind observation at 1000hPa. The observation is 1m.s^{-1} more westerly than the background. Left panel : using J_b (with the statistically calibrated $\mathcal{H}$ operator), right panel : using the analytical geostrophic balance as horizontal balance operator.

This is demonstrated in fig. 6 where the wind increments generated by the same equatorial wind observation are compared with and without a replacement of $\mathcal{H}$ by the analytical balance operator. The wind increments generated by a mass observation (not shown) are identical in both cases and they are completely non-rotational (the mass/wind coupling near the equator has nothing to do with the horizontal balance).

The conclusion is that the horizontal balance operator is essentially identical to the usual geostrophic balance between vorticity and mass. This shows that in its current implementation the statistical calibration technique of $\mathcal{H}$ produces results that agree completely with the conventional geostrophic theory. To make the horizontal balance more sophisticated it would be necessary to make it depend on the vertical coordinate and to seek a more general algebraic form (in order to allow for horizontal and vertical changes in the mass/wind coupling). This could be part of a future improvement of the J_b formulation.

Vertical balance operators

The vertical balance operators have a complex and noisy structure. The noise may be mere sampling noise (particularly at large scales), but it may also reflect the fact that no numerically stable linear regression could be found by the calibration algorithm. This is not as worrying as it sounds, because all that counts for the analysis is the final $\mathbf{B}$ covariance matrix. In this matrix, the balance operators are always multiplied by covariance matrices. The effective covariance structures (shown in another section) do not exhibit so much noise because the covariances are less noisy and their correlations effectively kill most of the sampling noise. Hence the noise is usually a consequence of the linear regression trying to extract information from mutually

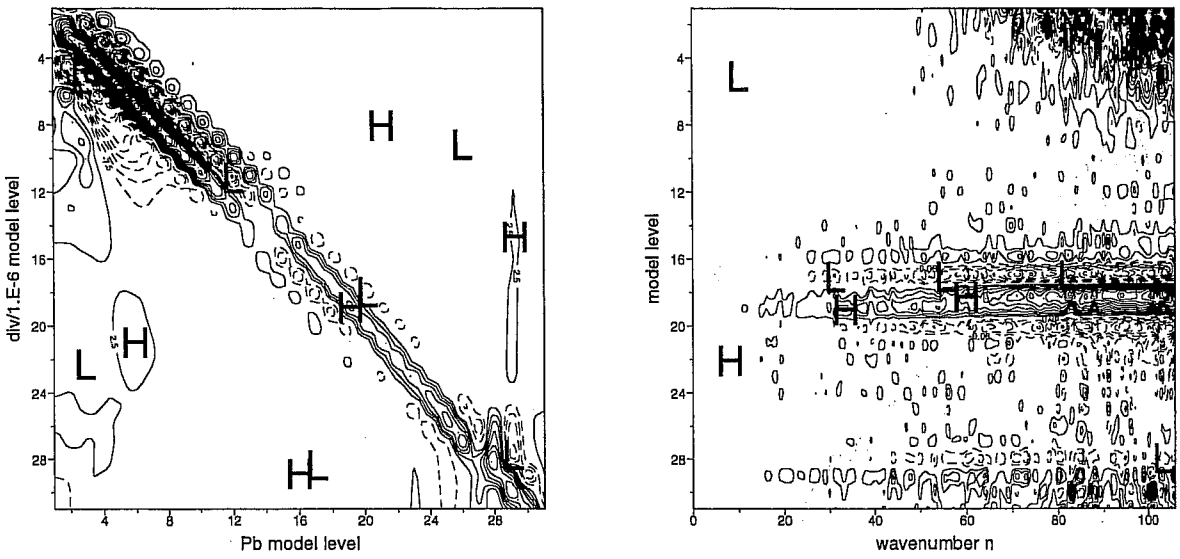


Figure 7: Representation of the coefficients of the vertical $P_b \mapsto \eta_b$ balance operator $\mathcal{M}$ in spectral average (left panel) and in a spectral cross-section of its lines at model level 18 (right panel), i.e. it is the sensitivity of η_b at level 18 to the vertical profile of P_b. The plot scaling is arbitrary, but the isolines are the same in both panels.

correlated predictors. The actual performance of the balance calibration must be appreciated in terms of the amount of explained variance. The aim of the operator plots in this section is to give a flavour of what is inside the balance operators, and one should be very cautious in looking for a scientific interpretation.

The $\mathcal{M}$ operator links the balanced mass P_b with divergence. The spectrally averaged vertical operator is displayed in fig.7 which shows that the operator is concentrated around the diagonal, i.e. the balanced divergence is generated by vorticity-implied mass at or around the same level. The sign of the coupling, however, is not obvious because the operator changes its sign rapidly with level index. The operator structure is complex near the ground, and it is sensitive to the choice of the reference (T, p_s) values in the definition of the linearized mass variable (but the effective correlation structure is not). A cross-section of the actual coefficients in shown in fig.7 too ; as in the other operators presented below, there is a lot of noise, but the structure near the diagonal is more or less the same at all wavenumbers, only its amplitude changes, which reflects both the spectra of the variables at hand and their mutual correlations. The operator has much more amplitude at the small scales, reflecting the fact that divergence is a smaller-scale field than geopotential.

The $\mathcal{N}$ operator links P_b with (T, p_s). Its spectral average is shown in fig.8 for which the same remarks as in the previous paragraph can be made. A spectral cross-section of the operator would reveal the same kind of structures, with again a higher amplitude of the operator at small scales. The spiked structure of the average $P_b \mapsto p_s$ operator near the ground is an artifact of the large P_b vertical correlations in this region. One can only tell that a positive p_s increment will be caused by an increase in low-level geopotential, which makes physical sense.

The P operator links η_u with (T, p_s). It will be shown later in this paper that η_u can be identified with total divergence. The spectral average is shown in fig.9 . Only wavenumbers larger than 5 have been retained in the average because the operator is very big and noisy for

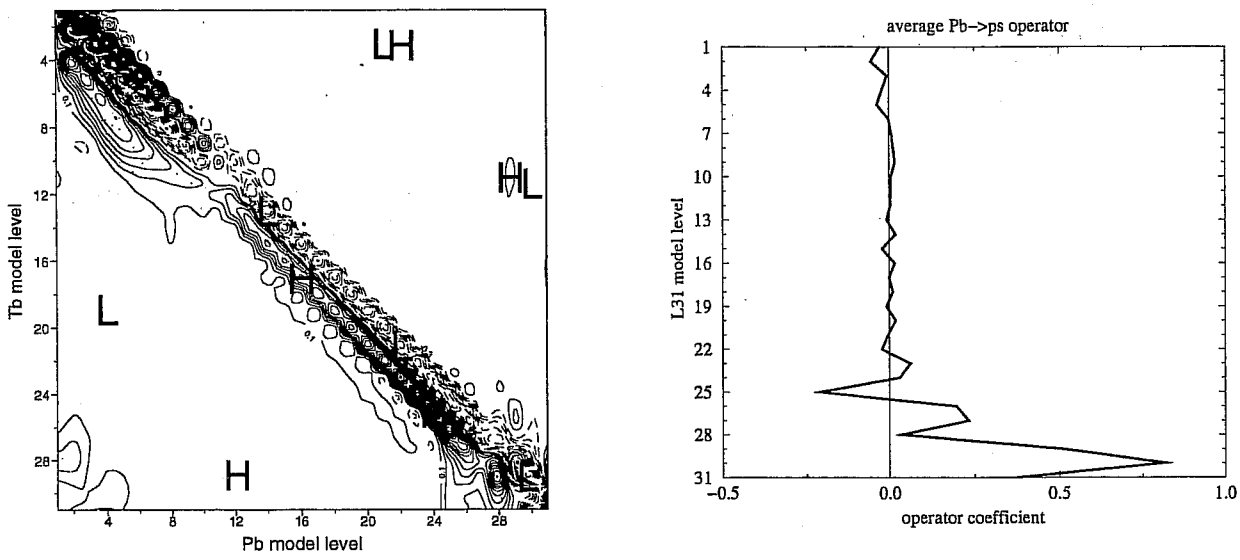


Figure 8: Representation of the coefficients of the vertical $P_b \mapsto (T, p_s)_b$ balance operator $\mathcal{N}$ in spectral average for the $P_b \mapsto T_b$ part (left panel) and for the $P_b \mapsto p_s$ part (right panel). The units are arbitrary.

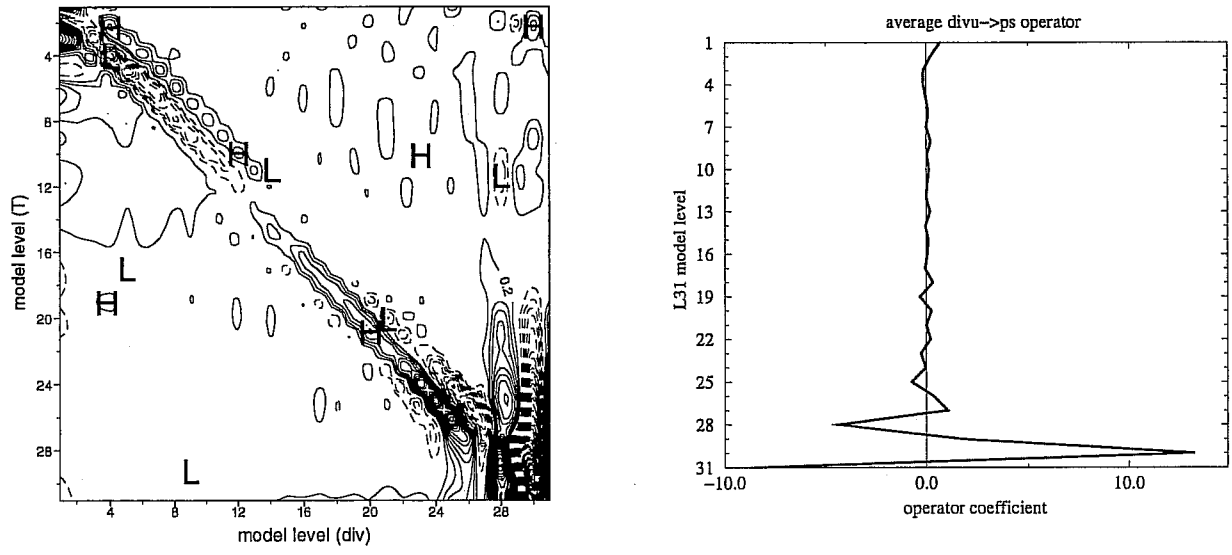


Figure 9: Representation of the coefficients of the vertical $P_b \mapsto (T, p_s)_b$ balance operator $\mathcal{N}$ in spectral average beyond wavenumber 5 for the $P_b \mapsto T_b$ part (left panel) and for the $P_b \mapsto p_s$ part (right panel). The units are arbitrary.

the larger scales. Since at these scales the error variance of divergence and (T, p_s) is small, they are not very significant meteorologically speaking (except perhaps for tidal waves). Again, the same remarks as for operator M can be made.

A physical interpretation of these operators will only be attempted after examination of the covariance structures implied by the balance operators *and* the error covariance matrices.

5.2 Statistical performance of the balance

The balance operator is calibrated using a linear regression principle. The performance of this operator can be measured objectively by diagnosing the amount of explained variance in the actual meteorological fields. It shows to what extent the assumed balance relationship really exists in nature, and it gives an indication of the optimality of the assumed algebraic form of the relationship. The rms distance between the balanced fields and the real ones is provided by the variance v_u of the residual, the unbalanced fields. We compare it to the total variance v_t by the following quantity called *amount of explained variance* :

$$1 - \frac{v_u}{v_t}$$

Here we use the same dataset to calibrate the balance and to measure its performance, which should not be too overoptimistic because the statistical sampling is good, except perhaps for the very large scales. The geographical distribution of the balance performance will be presented in another section. The actual relative size of the balanced and unbalanced terms in J_b is not just a property of the balance operator, it depends on the algebraic restrictions made on the covariance operators (their spatial homogeneity in particular) ; it will be shown in the section about the effective covariances.

The global amounts of explained variance are as follows :

parameter	explained variance
divergence	9.4%
temperature	62%
surface pressure	92%

The spectral and vertical distribution of these quantities is presented in figure 10. The divergence is weakly explained by vorticity at all scales, mainly near the ground at synoptic scales and near the tropopause at small scales. The (T, p_s) variables are better explained at synoptic than small scales (mainly by the geostrophic coupling). The synoptic-scale temperature is well explained at all levels except near the ground, obviously because of surface processes like friction. The small-scale temperature is well explained in the stratosphere (its total variance is very small there anyway).

5.3 Raw Covariances

The total divergence and (T, p_s) covariances are not explicitly used in J_b, but they provide an important reference because they do not depend on the design of the balance. Hence we are going to examine correlations and variances for total vorticity, divergence, temperature, surface

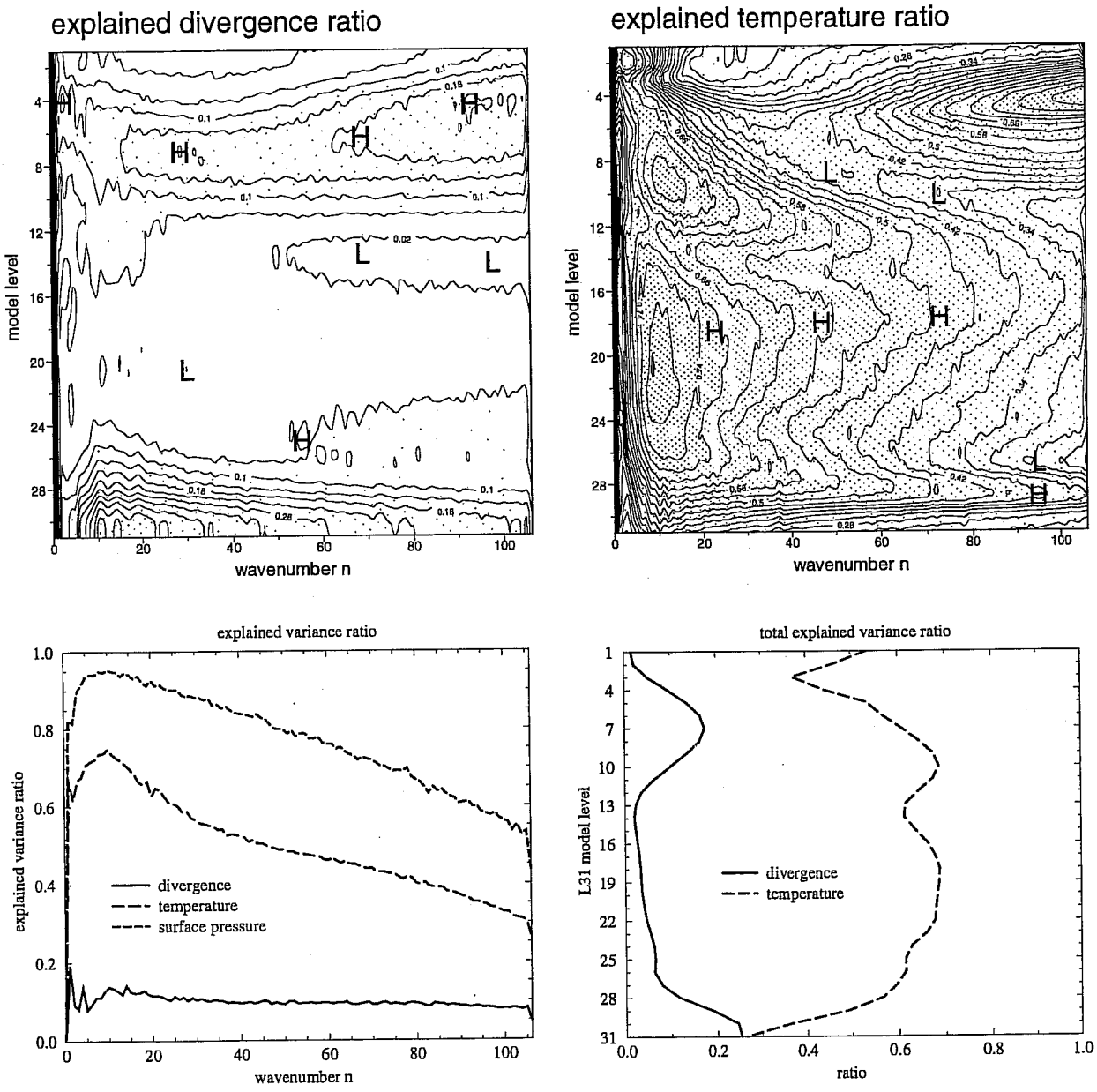


Figure 10: Distribution of the amount of explained variance (as defined in the text) for divergence (top left) temperature (top right), stratified according to total wavenumber (bottom left, including p_s), and to model level (bottom right).

pressure, specific humidity and unbalanced temperature and surface pressure. The unbalanced divergence is not plotted because it is virtually identical to the total divergence, except that the variance is slightly smaller ; this is because the balance explains only a very small amount of divergence errors.

For each variable we have a set of forecast difference vertical covariance matrices that are used as background error covariances modulo a global scaling of the variances (by 0.9) as explained earlier in this paper. The variances presented here are the unscaled ones. From each set of vertical covariance matrices, we can separate the distribution of variances as a function of level and wavenumber, and the vertical autocorrelation matrix for each wavenumber. The sum of all vertical covariance matrices yields a total covariance matrix for the whole spectrum, from which a correlation matrix can be extracted : it is an average of the correlations for each wavenumber weighted by the variance spectra, and it indicates the vertical increment structure we expect to find for isolated and localized observations of the same variable. The horizontal correlation structure in physical space can be diagnosed by a transform of the variance spectrum at the same level, following the theory explained in Courtier et al (1997) ; more precisely, the correlation structure as a function of distance $\rho(r)$ is derived from each variance spectrum $v(n)$ as follows :

1. the variance spectrum is normalized into a correlation power spectrum : $r(n) = v(n) / \sum_{i=0}^{\mathrm{NSMAX}} v(n)$,
2. the correlation power spectrum is converted into the *representation* of the correlation tensor (following the terminology of Courtier et al, 1997), by a multiplication by $\sqrt{2n+1}$ to account for the normalization of the Legendre polynomials, and a division by $2n+1$ to go from the power spectrum to the modal spectrum, i.e. $\rho(n) = r(n)/\sqrt{2n+1}$,
3. the correlation structure is computed on the collocation grid by a Legendre transform where the only non-zero components are for $m = 0$: $\rho(r) = \sum_{i=0}^{\mathrm{NSMAX}} P_n^0(r)\rho(n)$

In this section we present horizontal correlation structures at truncation T106 (on a 300–point meridian) ; the actual structures used in the ECMWF analysis are slightly smoother near the origin because the analysis is performed at T63 only.

Total variances

The vertical distribution of total standard errors is used to specify the horizontal averages of the background errors in the variational analysis. It is plotted in fig.11 for all variables. They all have small variations in the vertical except ζ which has a maximum in the upper troposphere (more or less at the midlatitude jet levels), and q which has completely different orders of magnitude in the stratosphere and in the lower troposphere. The large variation in the size of q is a cause for concern ; this, its boundedness and the non-homogeneity of its statistics as a function of background humidity make it a very ill-suited variable for a linear analysis scheme such as 3D-Var. Some work is needed to improve the situation in the future, starting with a better choice of analysis variable.

It will be shown in another section that the statistics of ζ, η and total (T, p_s) are dominated by error patterns in the midlatitudes. The unbalanced (T, p_s) statistics are built using patterns from the whole globe, and they are mainly important for the tropical part of the analysis. The

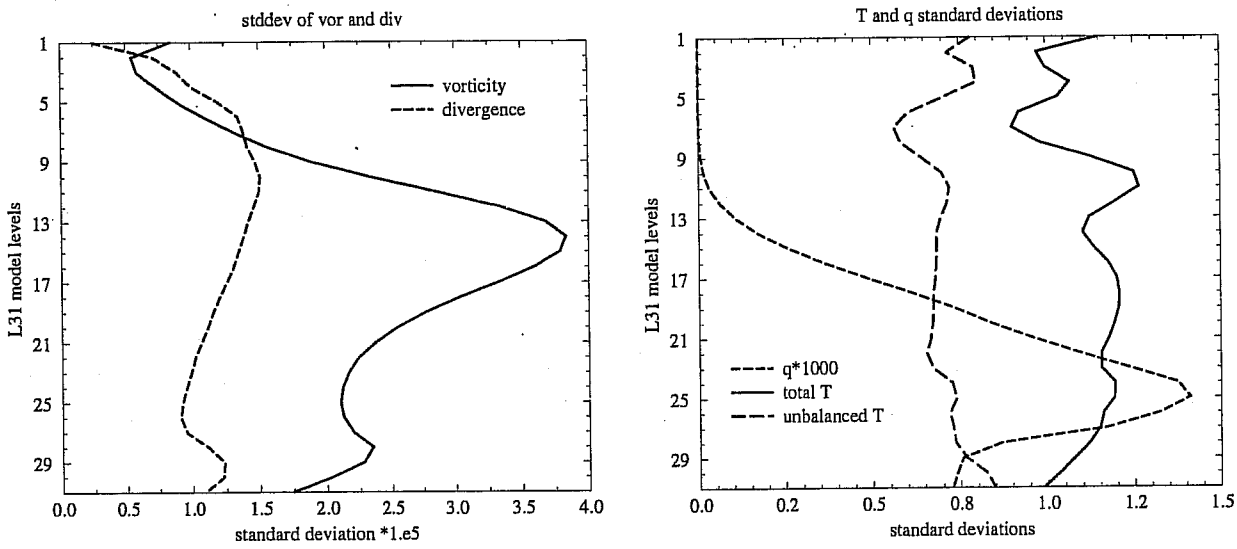


Figure 11: Vertical background standard error distribution in the unscaled error covariances. Left panel : vorticity and divergence multiplied by 10^5. Right panel : q multiplied by 1000, total temperature in the extratropics and unbalanced (i.e. tropical) temperature. The total and unbalanced p_s standard deviations are, respectively, 2hPa and 0.62hPa. J_b uses these values scaled by 0.9 .

spatial distribution of the errors in ζ and in the balanced variables is approximately modelized by the cycling algorithm which generates time-dependent geographical patterns like the one in fig.2. The cycling algorithm (see Fisher and Courtier 1995) uses the past history of the observing network and the climatological variances of the fields in order to estimate the spatial distribution of the short-range forecast errors in the assimilation system.

Vorticity

The vorticity covariances are depicted in fig.12 . The vorticity average vertical correlations are usually broad except in the stratosphere. They are even broader near the ground, which is a consequence of the vertical diffusion in the model. The sharpness of the vertical correlations increases with horizontal wavenumber, except at the very large scales ; this is commonly called "non-separability", and it is the sign of a kind of three-dimensional isotropy in the vorticity error patterns. Most of the variance of vorticity error comes from sub-synoptic scales, between wavenumbers 20 and 60, and from the top of the troposphere (the midlatitude jet levels). The horizontal correlation are sharp, and they are almost zero beyond 200km. They get broader with altitude in the stratosphere.

Divergence

The divergence covariances are plotted in fig.13 . As explained earlier, they are virtually identical to the covariances of unbalanced divergence. The vertical correlations are much sharper than for vorticity, and they are almost separable. The variances are distributed rather homogeneously in the vertical, except at the top, and they come from smaller scales than for vorticity. Accordingly, the horizontal correlations are sharper than for vorticity.

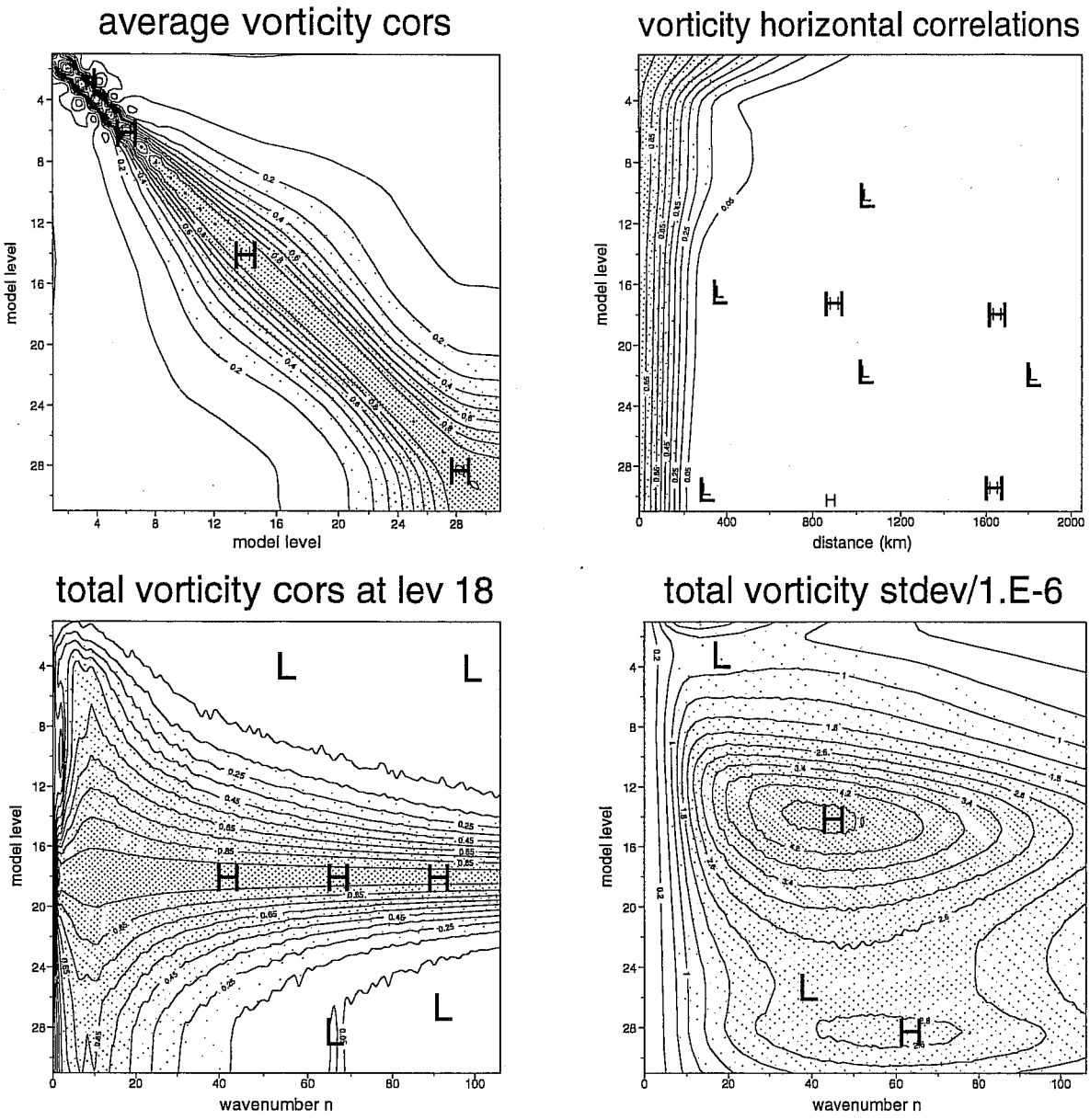


Figure 12: Covariances of vorticity ζ : average vertical correlations (top left panel), radial profiles of horizontal correlations (top right), vertical correlations with model level 18 i.e. 500hPa (bottom left), complete distribution of the standard deviations (bottom right). The correlation isolines values are 0.1 apart, with an isoline at 0.05, and the negative values are plotted with dashed isolines .

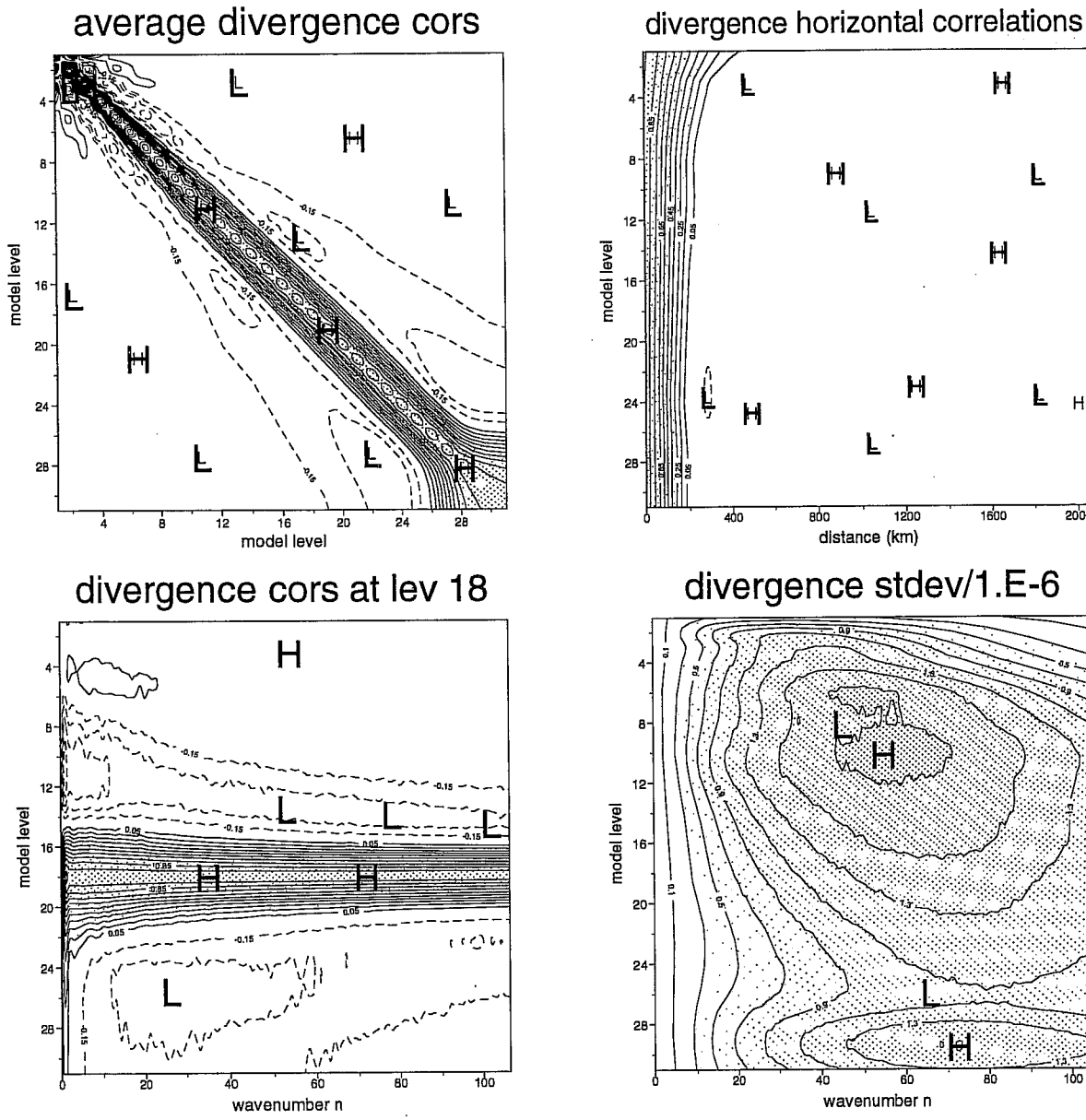


Figure 13: Covariances of divergence η, plotted like in fig.12 .

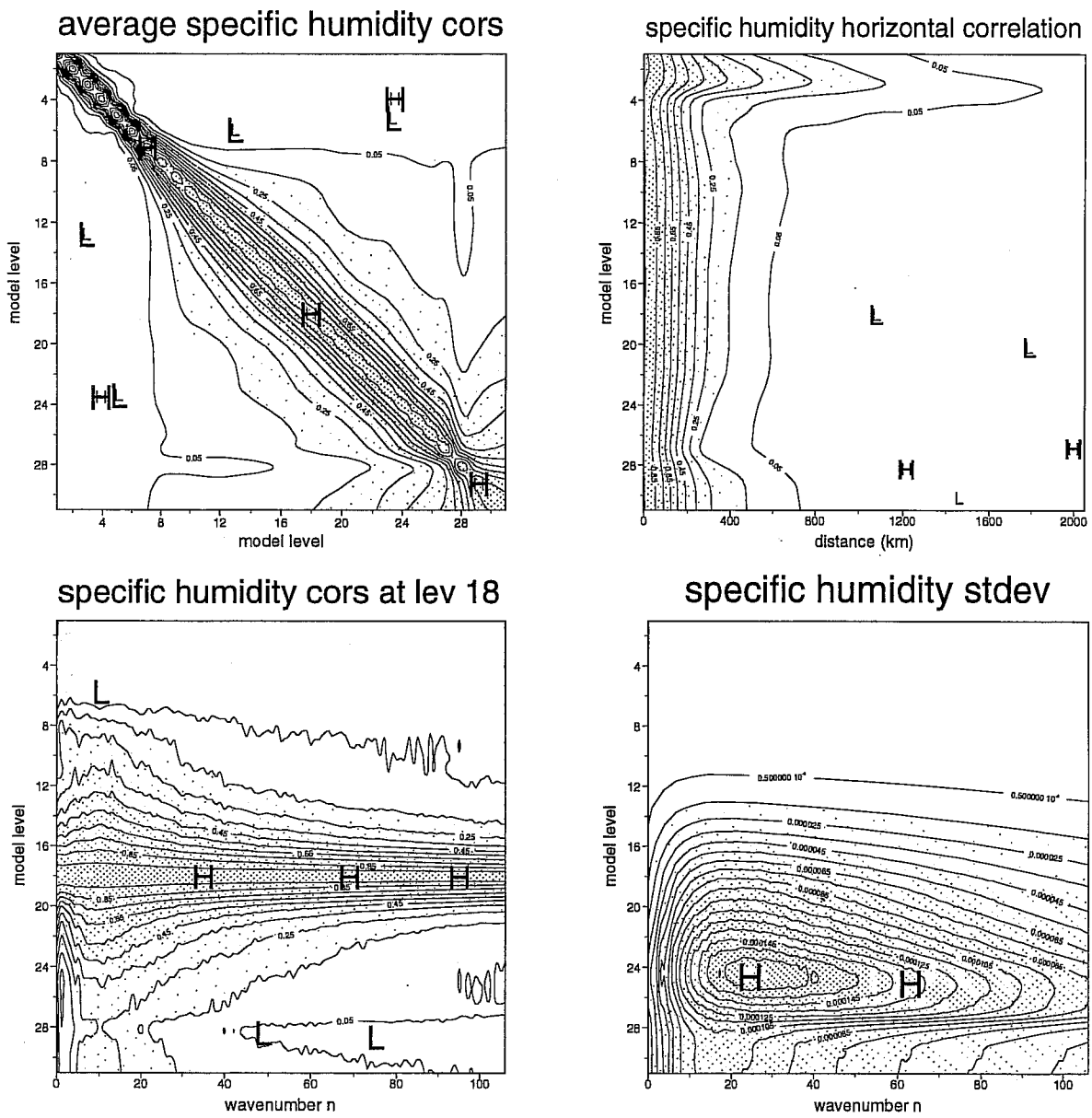


Figure 14: Covariances of specific humidity q, plotted like in fig.12 .

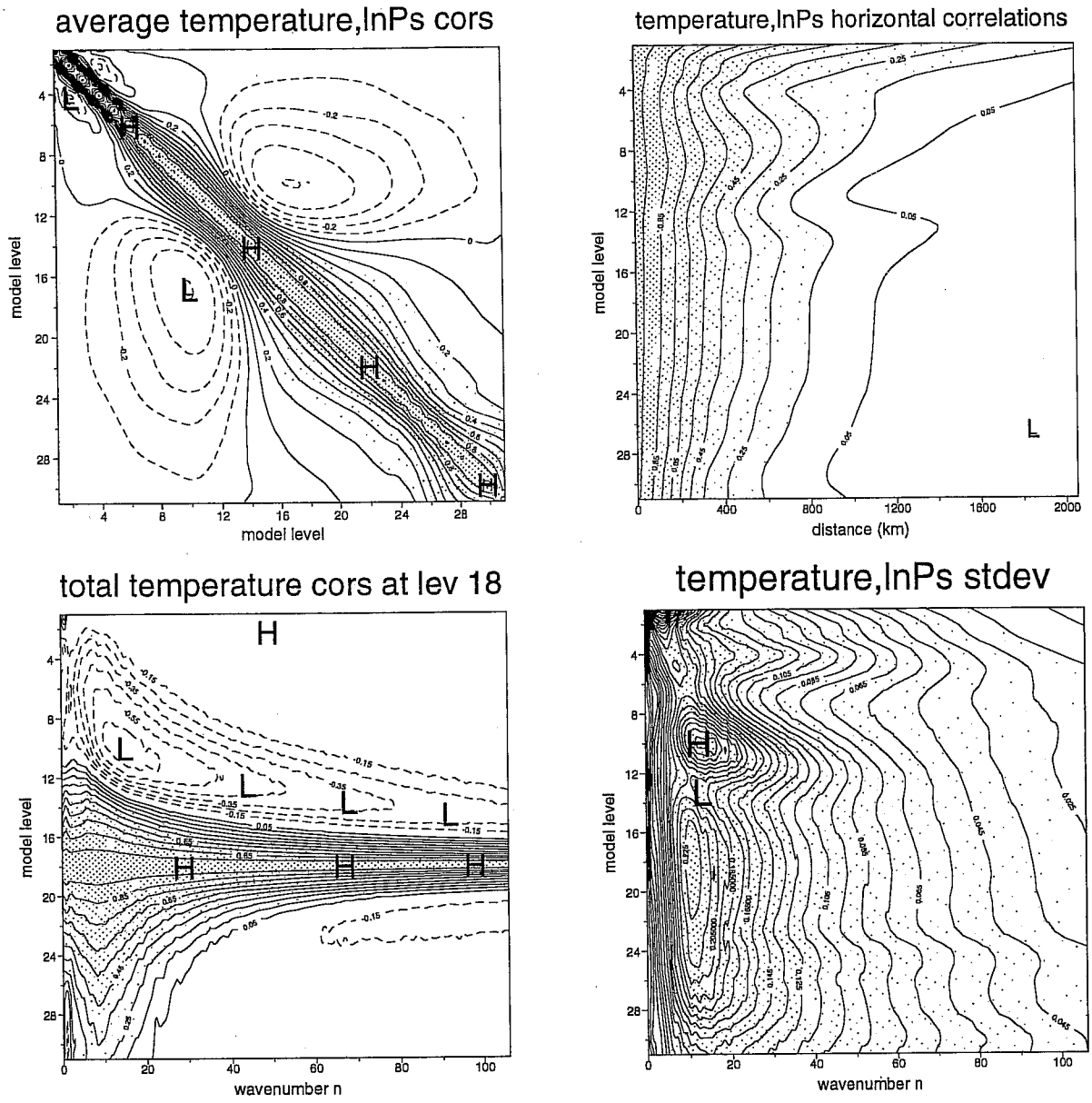


Figure 15: Autocovariances of total temperature T, plotted as in fig.12 .

Specific humidity

The humidity covariances in fig.14 are sharp in the vertical and fairly broad in the horizontal, with correlations of about 0.1 at 800km . The stratospheric structures are subject to caution because of the small variances at these levels. In the troposphere, the vertical correlations are strangely sharp for model level 28 (approximately 600m), and the variances decrease abruptly below this level ; this looks like an effect of the physical parameterizations used in the forecast model. The correlations are significantly non-separable. The bulk of the variance is located around level 24 (800hPa) and at synoptic and subsynoptic scales, roughly between wavenumbers 15 and 80, which suggests that the ECMWF analysis of even the average humidity structures would benefit from running 3D/4D-Var at a truncation higher than T63.

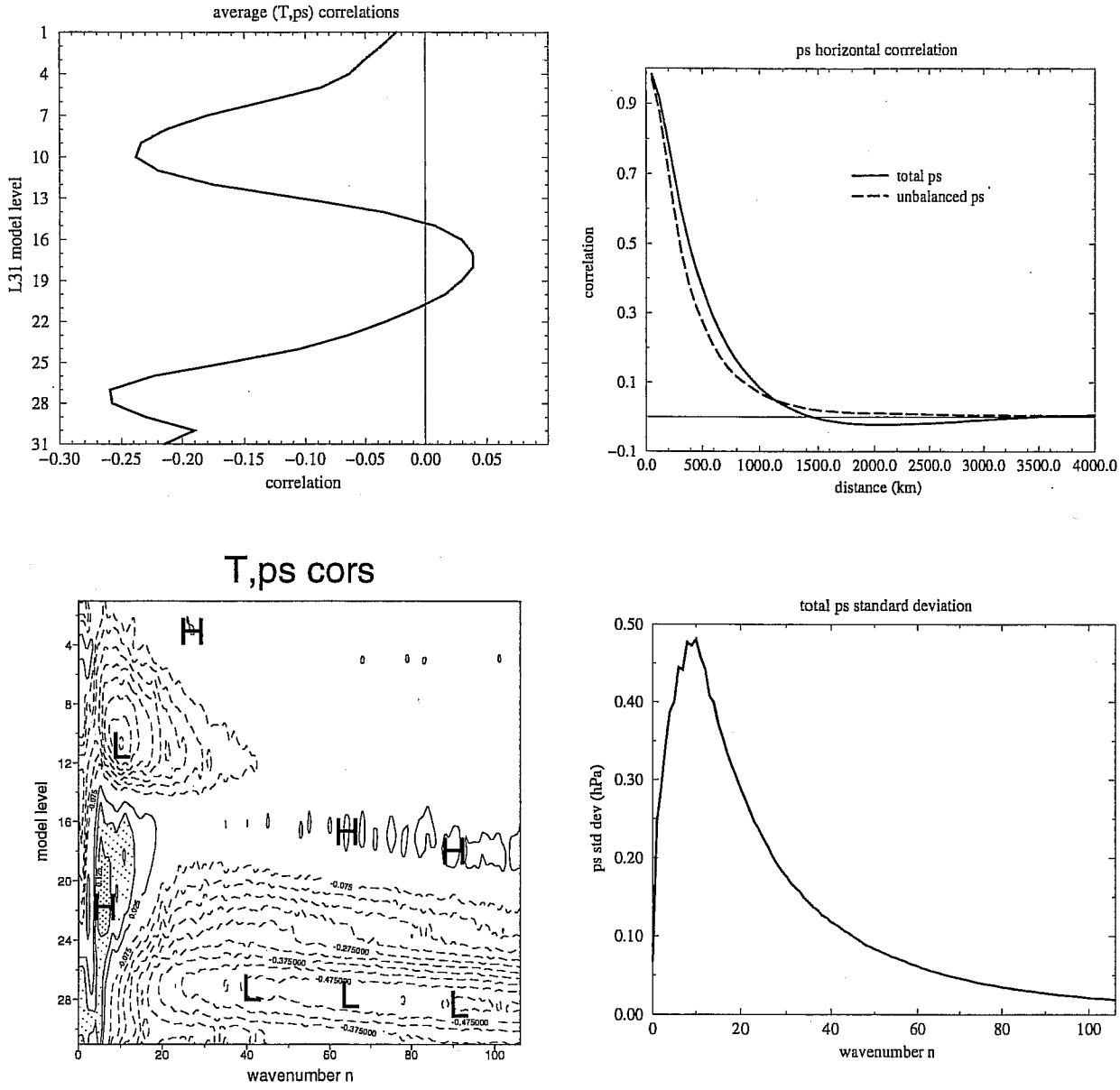


Figure 16: total (T, p_s) covariances related to surface pressure p_s : average vertical (T, p_s) correlations (top left panel), radial profile of horizontal p_s autocorrelation (top right, with the unbalanced p_s correlation for comparison), complete distribution of the (T, p_s) correlations (bottom left, the isolines values are 0.1 apart) and the p_s standard deviation spectrum (bottom right, in Pa).

Total temperature and surface pressure

The total (T, p_s) covariances are displayed in fig.15 and 16. The vertical T correlations are very broad in the troposphere and in the lower stratosphere, with negative correlations between the two regions (the boundary is around level 12, i.e. 270hPa). There is a substantial non-separability, and the broadness is mainly caused by large horizontal scale patterns between wavenumbers 5 and 20, with a large variance at all levels. Both T and p_s have a very small average variance beyond wavenumber 60 (remember these are global statistics, substantial smaller-scale (T, p_s) errors are known to exist in frontal regions, for instance). Accordingly, the horizontal T and p_s correlations are both broad ; the p_s horizontal correlation has a weakly negative lobe at a distance of 2000km. The (T, p_s) cross–correlations are generally negative, with about -0.25 correlations in the lower troposphere (at all scales except the very large ones) and around the tropopause (at large scales only). There is almost no (T, p_s) correlation between levels 13 and 24, i.e. between 300 and 750hPa.

These (T, p_s) statistics are representative of the midlatitudes. We shall see below that both the unbalanced and the tropical (T, p_s) variances and correlations are very different from the above picture.

Unbalanced temperature and surface pressure

The unbalanced (T, p_s) covariances are displayed in fig.17 and 18. The variances are smaller than for the total variables. The $(T, p_s)_u$ correlations are different from the total (T, p_s) ones : the vertical T correlations are much sharper, with less anticorrelation, and the change comes mainly from wavenumbers below 20, which is where most of the variance is explained by the balance operator M. The (T, p_s) cross correlations change too, with the disappearance of the negative correlation with tropospheric temperature ; it means that these correlated structures are geostrophic and well explained by the balance with vorticity, whereas the (T, p_s) correlated structures in the lower troposphere are incompatible with the balance, and they are left in the unbalanced statistics, with a negative correlation of almost -0.4 now. The horizontal unbalanced correlations are just a little bit sharper than those for the total variables.

5.4 Implied covariances

Despite the algebraic complexity of the implied **B** error covariance matrix, it is possible to illustrate rather simply the implied covariance operators in meteorological terms. One needs first to understand the relative weights of the balanced and unbalanced terms in **B**.

Amplitude of the balanced terms

In theory, the algebraic form of **B** makes it difficult to characterize the covariances of balanced η and (T, p_s), the autocovariances are equal to $MC_\zeta M^{\mathrm{T}}$ and $NC_\zeta N^{\mathrm{T}} + PC_{\eta_u} P^{\mathrm{T}}$ where M and N comprise simultaneously horizontal and vertical operators. Fortunately, we have seen in a previous section that the horizontal operator $\mathcal{H}$ is actually very similar to the analytical geostrophic balance, which is itself closely approximated by the f-plane balance relationship

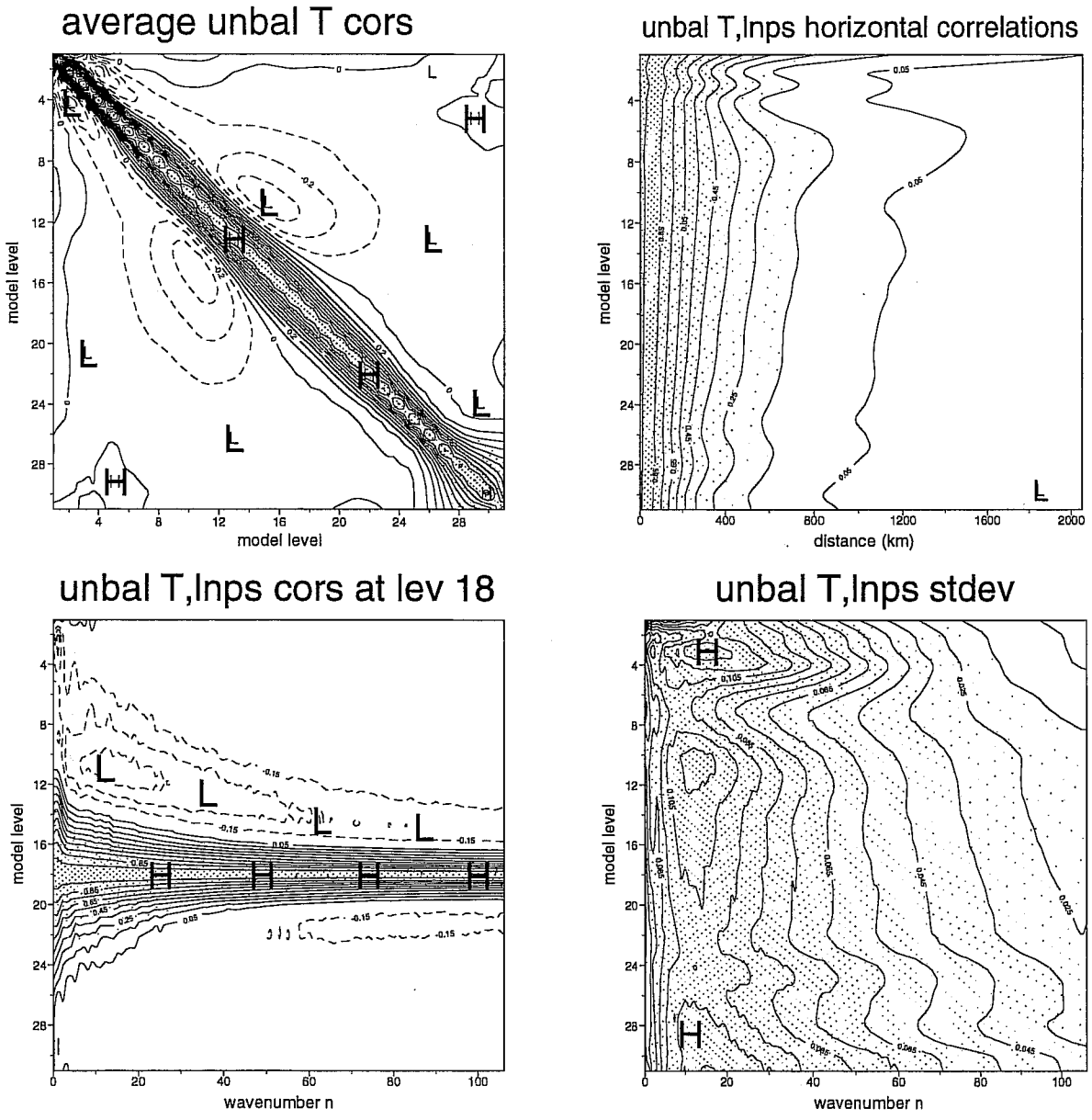


Figure 17: Autocovariances of unbalanced temperature T, plotted as in fig.12 .

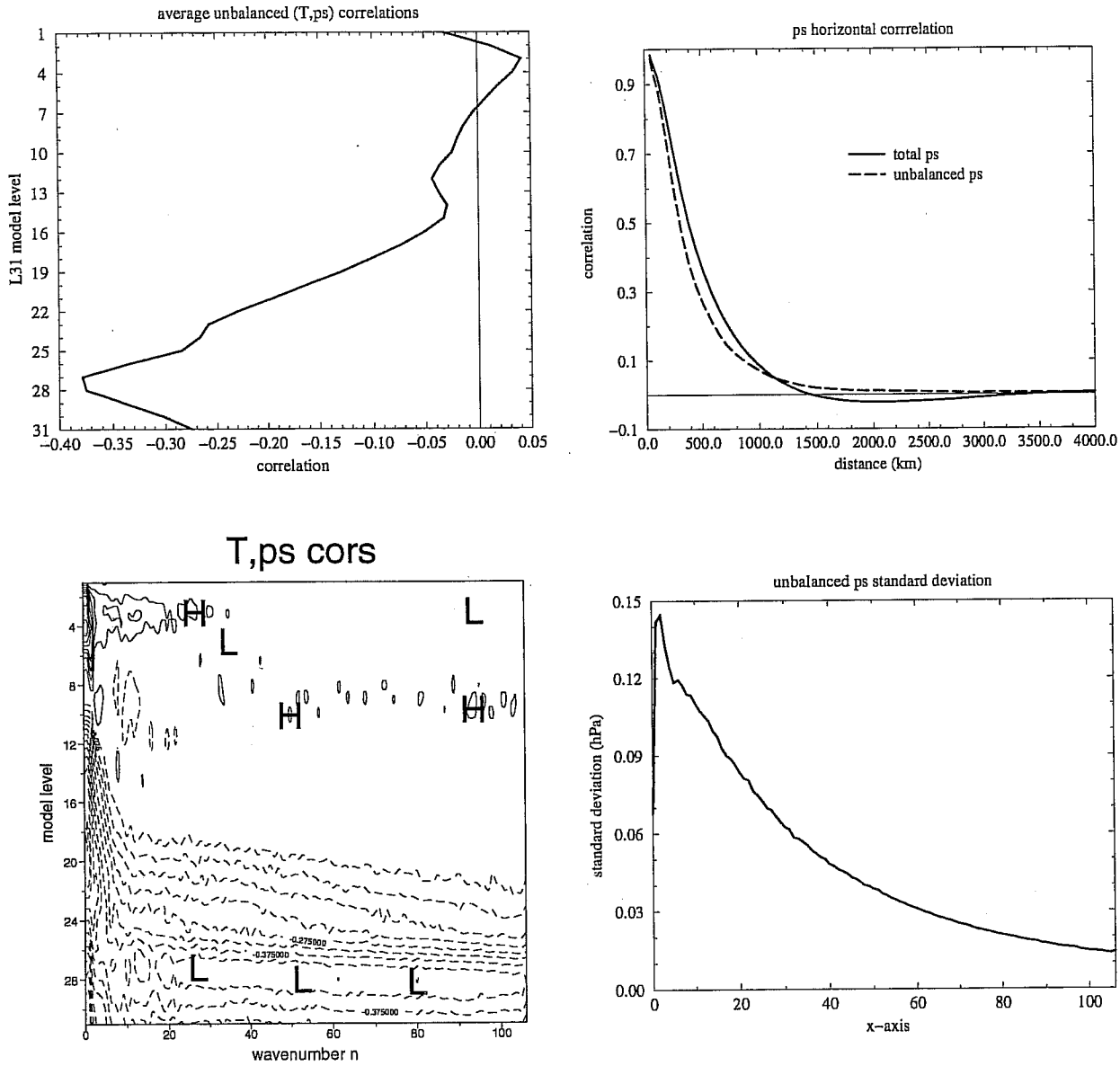


Figure 18: Unbalanced (T, p_s) covariances related to unbalanced surface pressure, plotted as in fig.16 .

$P_b = f\Delta^{-1}\zeta$. Hence the covariance matrix of balanced geopotential

$$C_{P_b} = \mathcal{H} C_\zeta \mathcal{H}^{\mathrm{T}}$$

is, to a very good approximation (except perhaps for the very large scales and for some β-effects near the equator), locally equal to the covariances of the inverse laplacian of vorticity (i.e. the streamfunction) times the Coriolis factor at the latitude φ considered, $f = 2\Omega \sin\varphi$:

$$C_{Pb} \simeq \hat{C}_{Pb}(\varphi) = f^2 \Delta^{-2} C_\zeta$$

In the tropical regions, f is small, so C_{P_b} is arguably negligible compared to the other covariance matrices, and the effective $\mathbf{B}$ matrix is :

$$\mathbf{B}_{tropics} = \begin{bmatrix} C_\zeta & 0 & 0 \\ 0 & C_{\eta_u} & C_{\eta_u} P^{\mathrm{T}} \\ 0 & P C_{\eta_u} & P C_{\eta_u} P^{\mathrm{T}} + C_{(T,p_s)_u} \end{bmatrix}$$

On a narrow band around any given extratropical latitude φ, we are going to assume that $\mathcal{H} C_\zeta \mathcal{H}^{\mathrm{T}} \simeq \hat{C}_{Pb}(\varphi)$, so that the effective extratropical $\mathbf{B}$ matrix is :

$$\mathbf{B}(\varphi) \simeq \begin{bmatrix} C_\zeta & \hat{C}_{Pb}\mathcal{M}^{\mathrm{T}} & \hat{C}_{Pb}\mathcal{N}^{\mathrm{T}} \\ \mathcal{M}\hat{C}_{Pb} & \mathcal{M}\hat{C}_{Pb}\mathcal{M}^{\mathrm{T}} + C_{\eta_u} & \mathcal{M}\hat{C}_{Pb}N^{\mathrm{T}} + C_{\eta_u}P^{\mathrm{T}} \\ \mathcal{N}\hat{C}_{Pb} & \mathcal{M}\hat{C}_{Pb}\mathcal{M}^{\mathrm{T}} + P C_{\eta_u} & \mathcal{N}\hat{C}_{Pb}\mathcal{N}^{\mathrm{T}} + P C_{\eta_u} P^{\mathrm{T}} + C_{(T,p_s)_u} \end{bmatrix}$$

The advantage of this expression is that each term is now the product of block-diagonal matrices (one block per set of vertical levels, for each total wavenumber), so that the corresponding non-separable covariance models can be calculated explicitly. All terms are constant, except the ones that depend on $\hat{C}_{P_b}$ which are modulated by the square of the sine of latitude according to $\hat{C}_{Pb}(\varphi) = f^2\Delta^{-2}C_\zeta$.

We are going to diagnose the balanced autocovariance terms using the following terminology : $\mathcal{M}\hat{C}_{Pb}\mathcal{M}^{\mathrm{T}}$ are the covariances of *balanced divergence*, $\mathcal{N}\hat{C}_{Pb}\mathcal{N}^{\mathrm{T}}$ are the covariances of *vor-balanced* (T, p_s), $P C_{\eta_u} P^{\mathrm{T}}$ are the covariances of *div-balanced* (T, p_s).

The relative importance of the balanced and unbalanced terms is summarized in fig.19 for two selected latitudes : the equator and 50N (or 50S). The ratio for the vor-balanced terms is exactly zero at the equator. The ratios are given by the variances of each balanced term divided by the variances of the sums of the balanced and unbalanced terms, i.e. the diagonal blocks in $\mathbf{B}$. We use variance ratios because they are expected to correspond to the ratios in the analysis increments themselves : roughly speaking, they indicate the amount of the increments that will be balanced. The scale dependence of the ratios is shown in fig.20 . These plots are extremely important for understanding the meteorological behaviour of J_b.

Most of (T, p_s) is balanced with ζ in the extratropics, and it corresponds to the geostrophic balance. The surface pressure p_s is almost completely vor-balanced, T is 80% vor-balanced at most levels, but less (60%) in the mid-stratosphere and near the ground (less than 40%), obviously because of surface drag. The vor-balance on (T, p_s) is weaker for smaller horizontal scales, and the non-separability of the T covariances implies that it is also true of shallow T errors. This is consistent with what is known about the geostrophic mass/wind coupling in the midlatitudes. The other types of balance are more original.

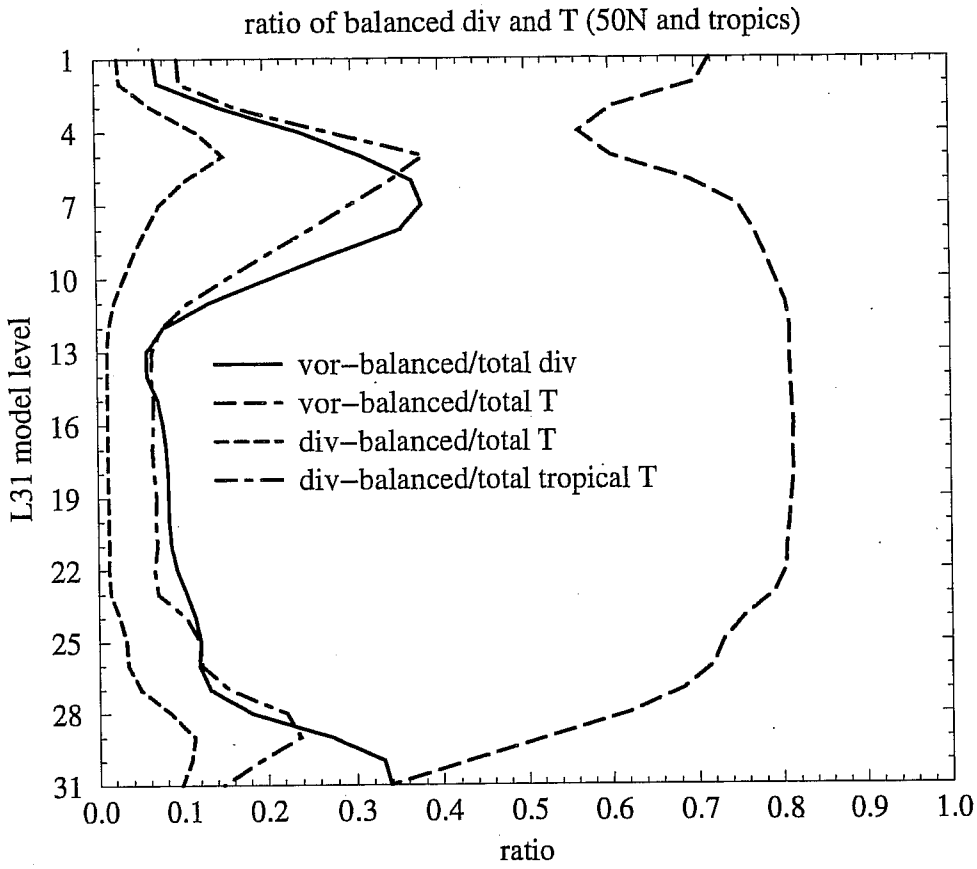


Figure 19: Vertical variations of the ratio between balanced and total covariance terms for divergence and temperature, plotted against model level. The 'tropical' curve is valid at the equator, the others are valid at 50N or 50S. For p_s, the div-balanced ratio is 0.02 at 50N and 0.18 on the equator, the vor-balanced ratio is 0.91 at 50N.

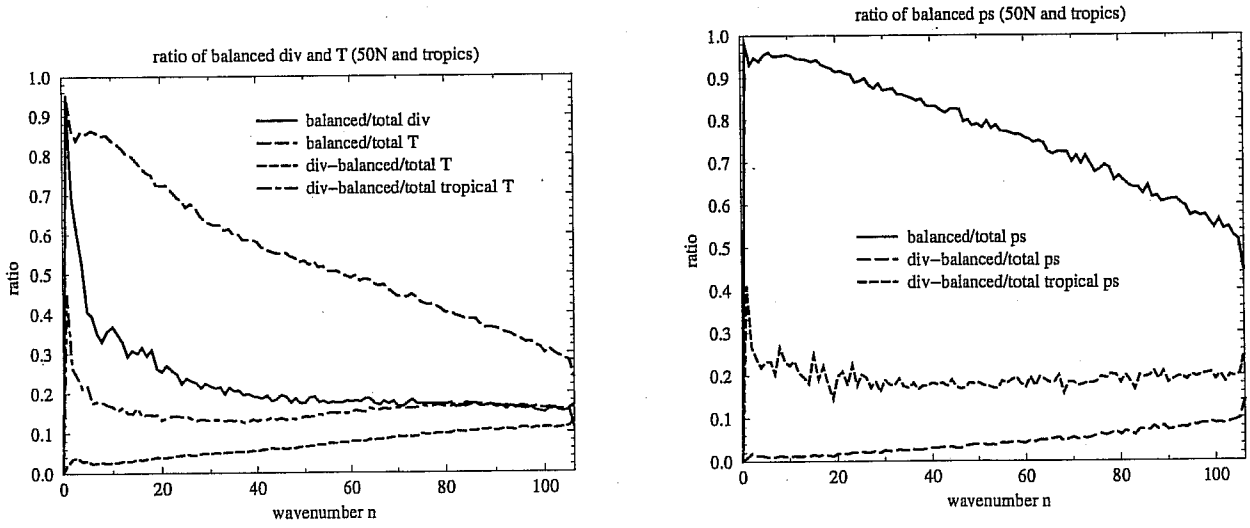


Figure 20: As fig.19, with the variance ratios plotted against total wavenumber.

In the midlatitudes, the div-balanced part of (T, p_s) is completely negligible, except perhaps for T in the mid-stratosphere and near the ground (only 10%, at small scales). However, near the equator it rises to significant values which affect all scales : more than 20% for p_s and for T in the mid-stratosphere and near the ground, about 10% at other levels. It means that the J_b formulation couples the tropical analysis of mass with divergent wind, which sounds very interesting for the assimilation of tropical convective systems and cyclones.

There is a direct link between divergence and vorticity, in the extratropics only. About 35% of divergence is vor-balanced in the lower troposphere and near the ground, 10% at the other levels. Since vorticity is strongly linked with (T, p_s) as well, this implies an indirect relationship between divergence and (T, p_s), too, which is likely to be similar to the one found above in the tropical regions. The coupling of divergence with the other variables is supposed to create more realistic ageostrophic circulations in the wind analysis at all scales (e.g. Ekman pumping near the ground in extratropical low-pressure systems). This is consistent with the decrease in the amount of balanced (T, p_s) found at these levels.

This qualitative analysis shows that some useful approximations can be made in interpreting the J_b covariances : in the extratropics, there is an important coupling between vorticity and (T, p_s), but only a weak coupling with divergence, which is predominantly uncoupled. Near the equator, the analysis is almost completely univariate, except for a small coupling between (T, p_s) and divergence, so that the effective tropical autocovariances are almost equal to the unbalanced η and (T, p_s) covariances already presented. To put it in a nutshell, the only parameter for which the autocovariances are significantly affected by the balance is the extratropical (T, p_s).

Implied (T, p_s) covariances

The effective tropical (T, p_s) autocovariances are not presented here, because they are practically identical to the unbalanced (T, p_s) covariances already shown in a previous section (there is just a small discrepancy in the variances which is documented in fig.21). This is because the div-balanced part of the tropical (T, p_s) covariances is small compared with the unbalanced part. The important effect of the div-balanced part is the implied mass/wind cross-correlation, not in the change of the (T, p_s) autocovariances.

The extratropical (T, p_s) covariances are presented in figure 21, which is valid on an f-plane at 50N (or 50S). The correlations are very different from the tropical ones and they are extremely close to the total correlations, which shows how good the balance is at reconstructing (T, p_s) covariances using vorticity covariances. The variances are compared with the tropical, global and unbalanced ones, which shows that the effective background error variances of (T, p_s) are much smaller in the tropics than in the extratropics. The correlation structures are close to the globally averaged ones. The variances at 50N are larger than the global values, which are more representative of the error amplitude statistics at, approximately, latitudes 40N and 40S.

Implied mass covariances

As explained earlier, the linearized mass variable P used to calibrate operator $\mathcal{H}$ is akin to a geopotential height (times g) ; its definition can be used to transform (T, p_s) vertical profiles

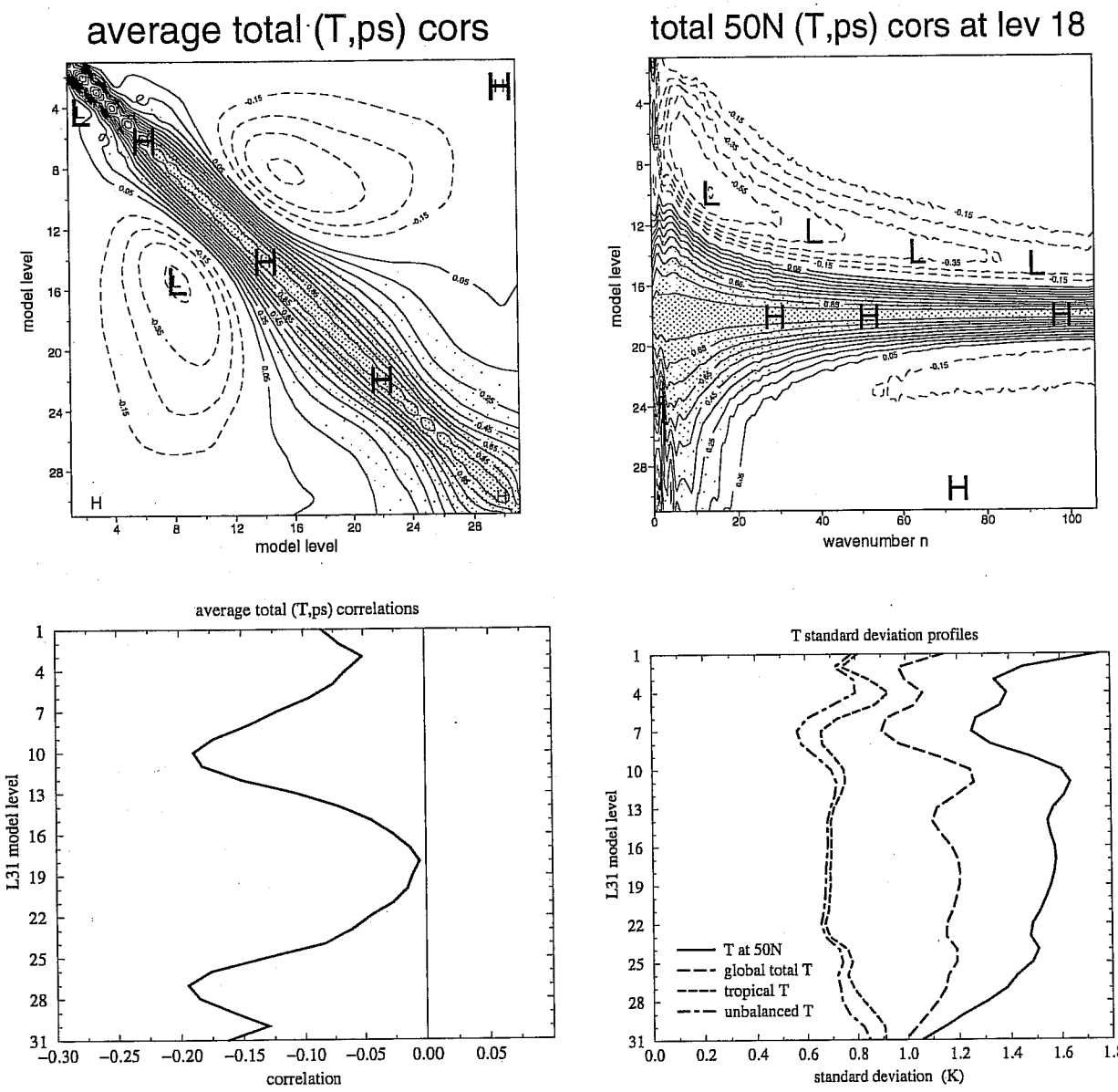


Figure 21: Effective (T, p_s) covariances on an f-plane at 50N or 50S : average (T, T) vertical correlations (top left), spectral dependence of the (T, T) vertical correlations with model level 18 i.e. 500hPa (top right), average (T, p_s) cross-correlations (bottom left), temperature standard deviation profile (bottom right) compared with the standard deviations of tropical, total (i.e. global) and unbalanced temperature. The corresponding p_s standard deviations are, respectively : 2.56, 0.88, 1.98 and 0.62hPa. The horizontal correlations are very similar (only slightly broader) to the ones for the total (T, p_s) already presented.

into P profiles using the following application $\mathcal{S}$ at each model level l :

$$\mathcal{S} : (T, p_s) \mapsto P$$

$$P_l = \sum_{i=\text{NFLEV}}^{l} RT_i \Delta \log p_i + RT_{ref} \log p_s$$

(in the IFS/Arpège code, $\mathcal{S}$ is the sigam operator) We can use this operator to diagnose the geopotential structure functions using $\mathcal{S}C_{(T,p_s)}\mathcal{S}^{\mathrm{T}}$ as approximations to the geopotential covariances in the tropics (using the unbalanced (T, p_s) covariances) and in the extratropics (using the total (T, p_s) covariances approximated on an f-plane). Figure 22 depicts them together with the "balanced geopotential" $C_{P_b} = \mathcal{H}C_\zeta\mathcal{H}^{\mathrm{T}}$ average covariance matrix.

The tropical and extratropical mass covariances are depicted in figure 22 ; Like for (T, p_s), they are very different from each other, with sharper horizontal and vertical correlations in the tropics. There is a substantial non-separability in the extratropics, but practically none near the equator (not shown). The correlation between the top and the bottom of the atmosphere is close to zero in the extratropics, and to 0.1 at the equator. The correlations of (global) total mass are not shown because they are almost exactly identical to the extratropical ones. The standard deviations (displayed in meters times the g constant) are much larger in the extratropics than in the tropics, and again the global statistics are representative of latitude 40N (or 40S) or so. The unbalanced mass statistics (not shown) are practically identical to those of tropical mass, both in terms of correlations and variances.

The C_{P_b} covariances are very different from the effective geopotential covariances (which are derived from complete (T, p_s) statistics instead of ζ statistics), both at the equator or in the extratropics. There is not just a scaling factor between the balanced and the extratropical mass standard deviations ; the C_{P_b} standard deviation are small near the ground, because the vorticity is reduced by surface friction in this region. This is an artifact of the constraint imposed on the design of the balance, and it is compensated by the $\mathcal{N}$ vertical balance operator which implies that the real geopotential variance is almost vertically constant near the ground, which makes more physical sense. The conclusion is that, although P_b is taken to be the linearized mass in the calibration of the balance operators, in J_b it is just an intermediate variable which is a poor approximation of the real geopotential height.

5.5 The cross-covariances

Using the same methodology as before it is possible to diagnose the cross-covariances, i.e. the off-diagonal blocks of matrix **B**, for several latitudes. The implied information is very rich and cannot be presented if the length of this paper is to be kept reasonable. Some concrete information about the cross-covariances will be shown in the section on simulated observations. Here we will just demonstrate on some very simple examples how the multivariate character of J_b can be traced back to the cross-covariance blocks of the **B** matrix.

Figure 23 shows the spectral averages of the MC_ζ operator linking ζ and η, and of the p_s part of the PC_{η_u} operator, i.e. the link between (unbalanced) divergence and surface pressure. The first one is only important in the extratropics, and it has been computed for latitude 50N here. It can be read like this : for an increment of vorticity at any given level, the corresponding column of the plot gives the implied vertical distribution of divergence. This in turn can be connected to mass increments through the geostrophic coupling. The plot could

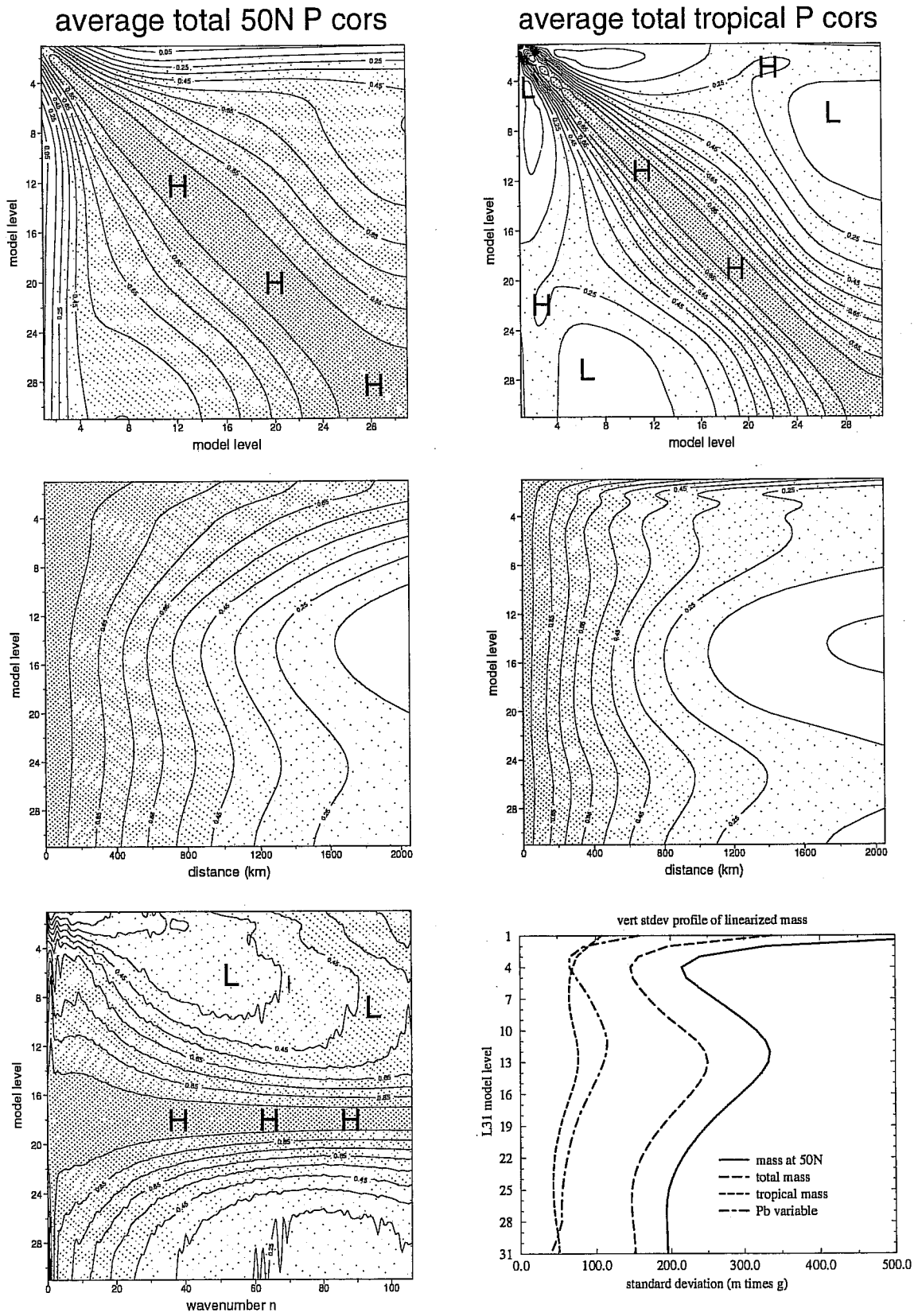


Figure 22: Effective covariances of linearized mass : vertical average correlations at 50N (top left), on the equator (top right), horizontal correlarions at 50N (middle left), on the equator (middle right), spectral dependence of the vertical correlation with level 18 i.e. 500hPa at 50N (bottom left), vertical standard deviation profiles compared with total mass and P_b variable (bottom right).

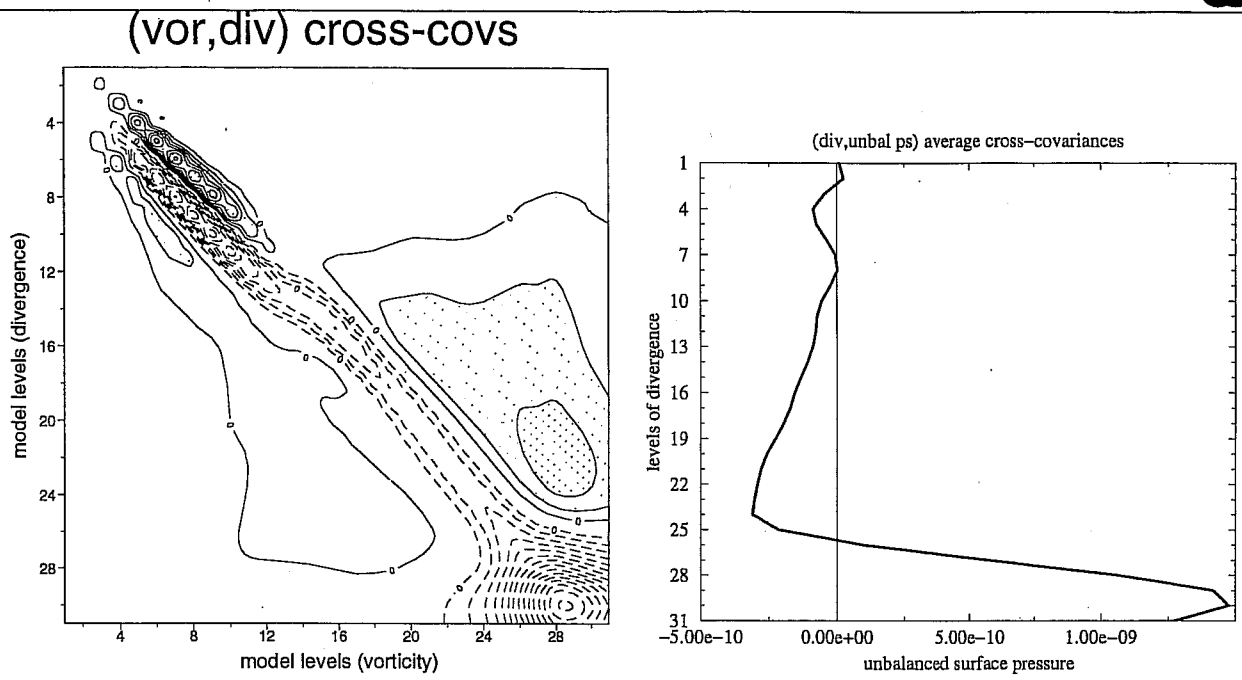


Figure 23: Spectral average of the cross-covariance operators between vorticity and divergence (left) and between unbalanced divergence and surface pressure (right).

also be read backwards, i.e. a divergence increment at a given level yields a vorticity profile given by the corresponding line. The operator has not been normalized by standard deviations, it shows cross-covariances. This is more meaningful than cross-correlations because there are vertical variations of the variances (e.g. in the upper stratosphere). The figure shows that a cyclonic wind increment near the ground will be associated with convergence near the ground, and divergence above level 25 (800hPa). A cyclonic increment at higher levels will only be associated with weak convergence at the same level. In the stratosphere the interpretation is more difficult because the possible ζ and η structures are constrained by the autocovariances of vorticity and divergence, so that the sharp lobes on the figure would actually be smoothed out.

The right-hand panel of figure 23 shows the average of the PC_{η_u} operator which connects divergence and surface pressure in the tropical regions (in the extratropics the "geostrophic" coupling term $NC_\zeta M^{\mathrm{T}}$ is likely to change the (η, p_s) cross-correlations). One sees that a negative surface pressure increment will result in convergent wind near the ground, and divergence above. The change of sign is at level 26, i.e. 1300m . It is tempting to interpret this as the signature of tropical convection.

6 Other diagnostics

6.1 Sampling issues

As explained before, it is absolutely necessary to use more independent error patterns than the number of levels in the J_b calibration procedure. It is desirable to use many more than that in order to have a correct statistical sampling, particularly at very large scales. The first J_b tests were carried out using 45-day samples (24/48-hour forecast differences taken one day apart from each other) during December 95/January 96, March/April 96, June/July 96,

September/October 96 and December 96/January 97. An extensive subjective comparison of the statistics between these 5 periods has been carried out, and it showed that there was no significant impact of the seasonal cycle, obviously because the J_b design does not distinguish between the Northern and Southern Hemispheres. However, there was a difference between the December/January 95 period and the other ones, because of the change in the ECMWF analysis system from OI to 3D-Var (at the end of January 1996). Consequently, it has been decided to neglect the seasonal effects and to recompute the statistics using one single 180-day homogeneous sample, leaving out the December/January 95 data. The 180-day statistics are the ones used in the operational J_b implementation of May 97, and in all the diagnostics presented in this paper.

Changing the sample size from 45 to 180 days altered only a couple of features in the statistics. There was a reduction of the noise in the diagnostic plots that are not averages, e.g. in the spectral dependence of the vertical correlations. Perhaps more importantly, there are some significant changes in the vertical correlation matrices for wavenumbers n between 0 and 3 (for higher wavenumbers the changes are not noticeable). Correspondingly, the differences between the analysis increments are mainly found in the very large scales. This may be important for the assimilation of tidal waves, although the impact has not been fully investigated.

A weakness of the J_b calibration procedure is that, even with 180 days of data, the sampling of the very large-scale statistics is very suspicious. The problem is likely to get worse when the number of levels in the model is increased, because a large set of forecasts at the right vertical resolution must be rerun before the actual J_b calibration can be done. In the future it will become necessary to develop either a method to extrapolate vertically the J_b statistics, or to use an independent source of information in order to calibrate the very large-scale part of the error covariances (e.g. a low-resolution Kalman filter).

6.2 Spatial homogeneity of the covariances

A major weakness of the previous J_b structure functions was their lack of geographical variability. The contrast between tropical and extratropical (T, p_s) structures has been apparently solved with the new formulation, but it remains to be seen whether any important geographical information is still missing. In this section we use the NMC method to calculate features of the forecast errors on specific domains. In doing so there is the danger of pushing the NMC method outside its domain of validity. In particular, using 24/48-hour forecast differences instead of 6-hour forecast errors is likely to hide the effects of the spatial inhomogeneity of the observing network. In the Northern Hemisphere there are some large differences in the observation density between America, East Asia or Europe, and the Atlantic and Pacific Oceans. How the J_b statistics should be modified to account for these differences is important for improving the performance of the assimilation and forecast system over (say) Europe, but it is beyond the scope of this paper : it would be necessary to use observation departures and some form of Kalman filtering. Here we only try to account for the zonally averaged differences of the atmospheric dynamics. We use NMC error patterns from September/October 1996 in this study.

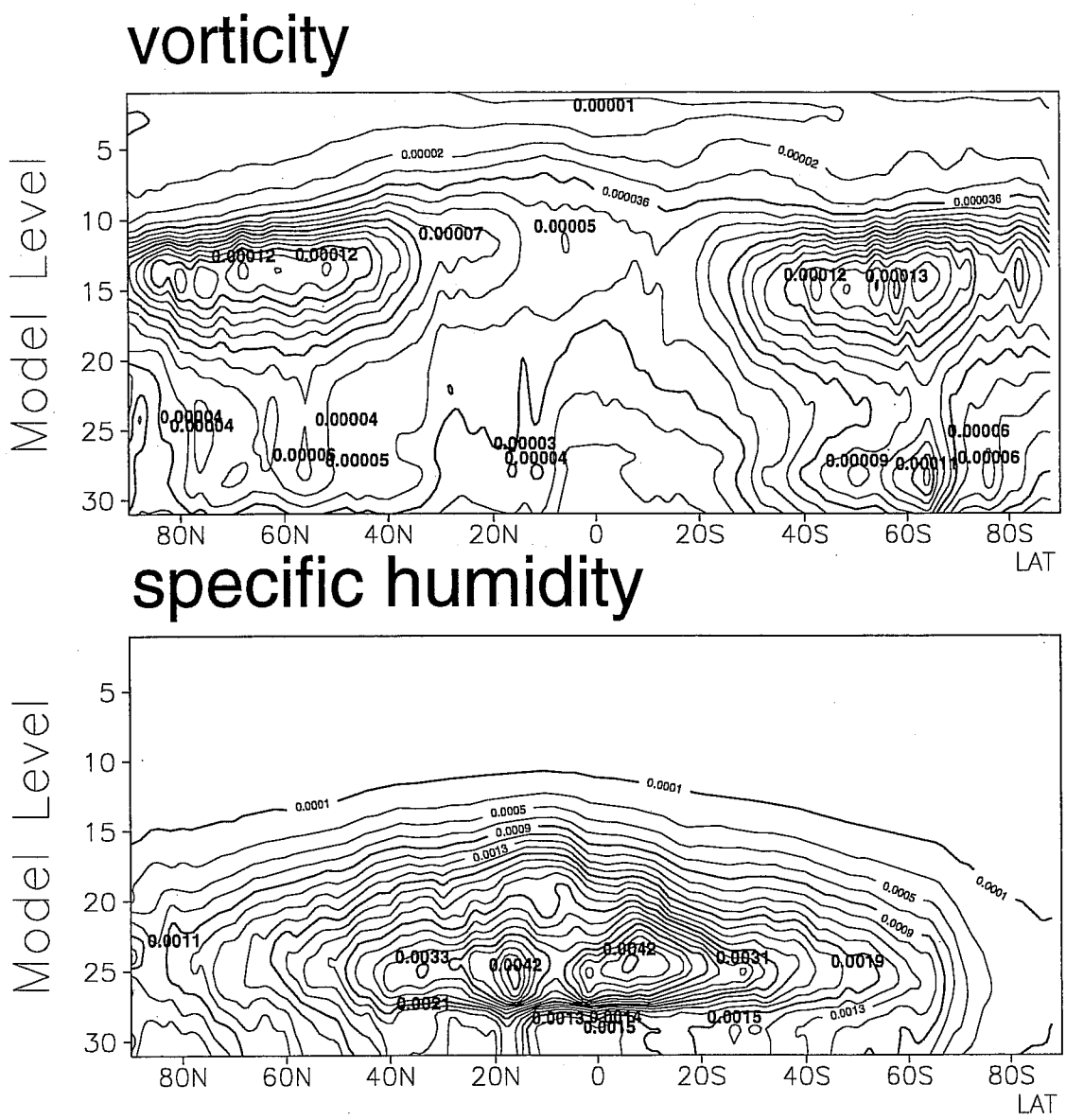


Figure 24: Meridional distribution of the total standard deviations of total ζ (top) and q (bottom).

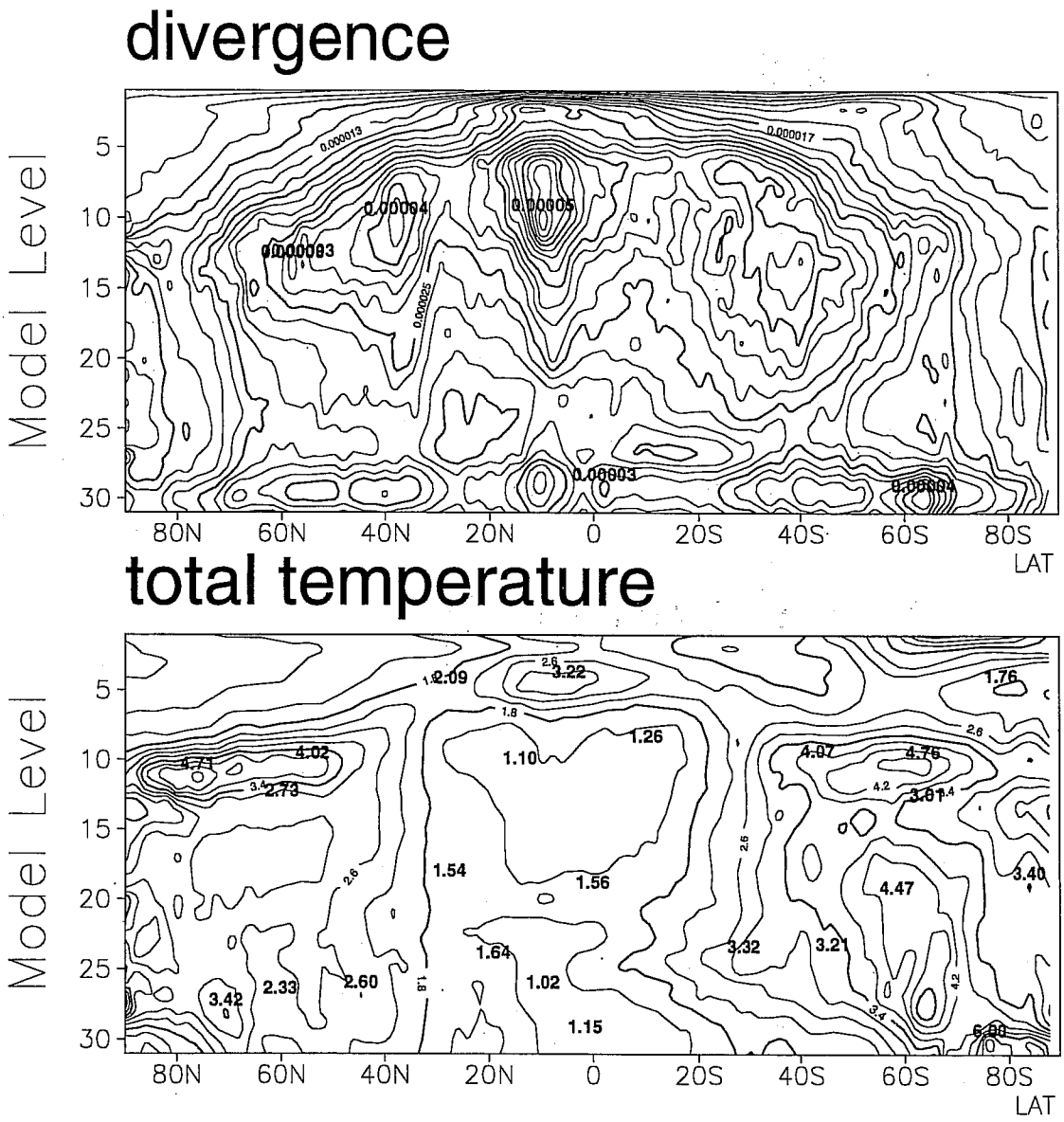


Figure 25: Meridional distribution of the total standard deviations of total η (top) and T (bottom). The total p_s standard deviation (not shown) is less than 1hPa between 25N and 25S, about 5hPa in the whole Northern Hemisphere, up to 10hPa in the Southern Hemisphere at 60S.

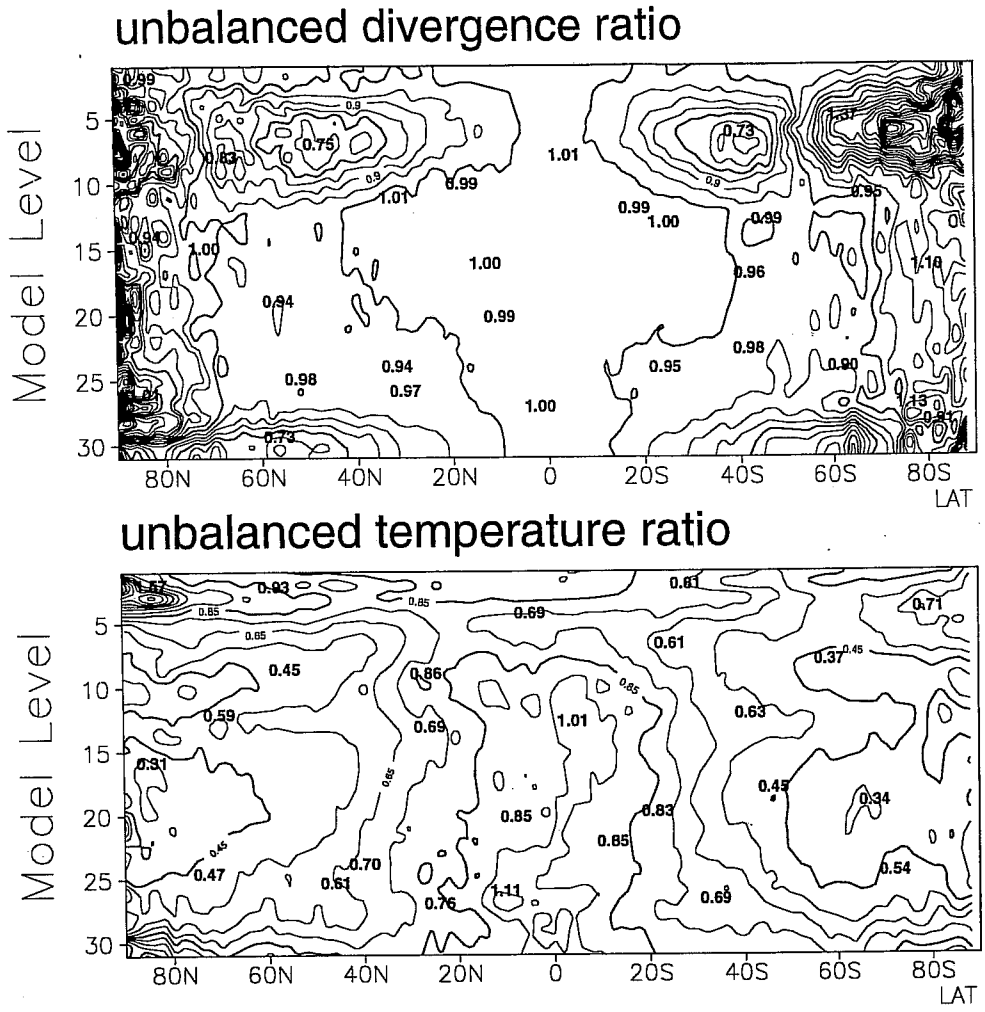


Figure 26: Meridional distribution of the ratios between the unbalanced and total fields of η (top) T (bottom). The unbalanced fields are constructed from the total ones using the inverse of the J_b balance operator. The ratios for p_s are 0.9 near the equator, 0.3 in the extratropical parts of both hemispheres.

Distribution of the variances

The maps of standard deviations of NMC forecast differences (not shown) are almost perfectly zonal : as expected, the NMC method hides the structure of the observing network. The meridional distribution of the standard deviations of the total variables is shown in figs.24 and 25 . The biases are small compared to the standard deviations, except between model levels 7 and 9 (between 130 and 200hPa) where temperature biases of 15% of the standard deviation are common, with a maximum in the tropical regions. The plots show the well-known importance of the midlatitude storm-tracks on wind and (T, p_s) near the tropopause and near the ground, and of the tropical convective systems on divergence and humidity.

The performance of the balance operator in geographical terms is shown in fig.26 : the higher the ratio, the lower the amount of explained variance. The graphs are noisy near the poles because of the poor sampling there. The balance with vorticity is only able to explain some divergence patterns in the extratropics near the ground and near the troppopause (the ratio can be greater than one if the balance tends to work the wrong way), which confirms the

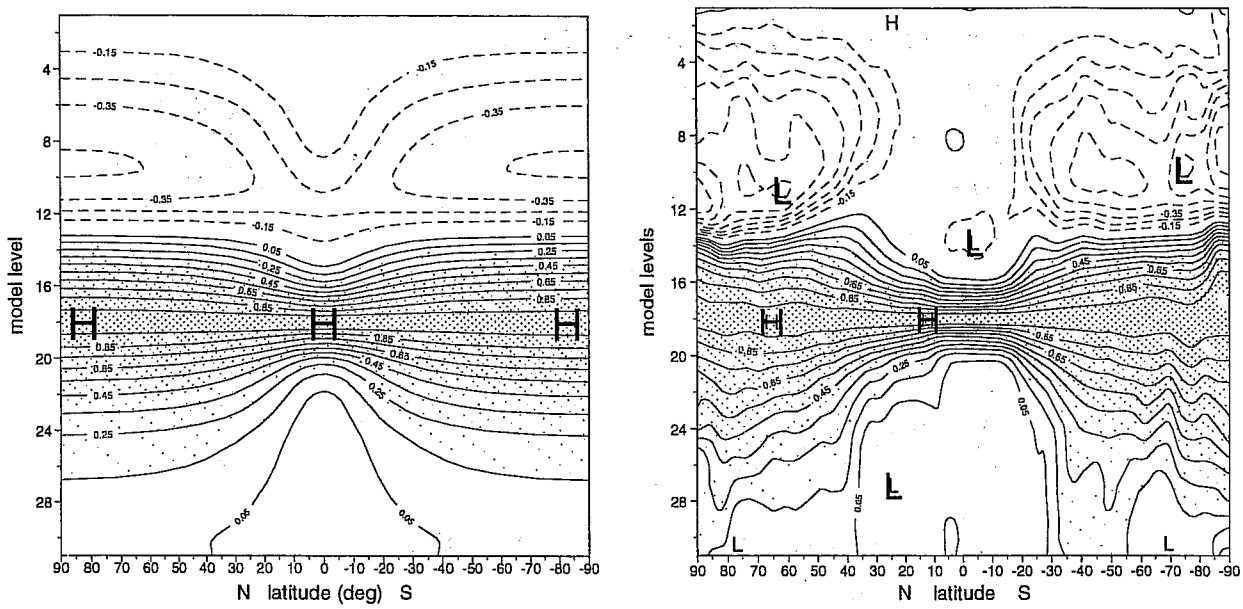


Figure 27: North-South variation of the total (T,T) vertical correlations at model level 18, in J_b (left panel), and in the NMC statistics (right panel).

diagnostics made earlier with the spectral covariances. The balanced temperature in the tropics only works near the tropopause, with 31% of variance explained locally (because of the link with divergence). In the extratropics, the balance works is more and more important nearer the poles, except near the ground, particularly in the Northern Hemisphere. This is obviously an effect of the surface friction.

The distribution of the variance of the unbalanced variables is not shown ; for divergence it is almost the same as for the total variable. For T it is a rather uniform distribution, because most the difference between the tropical and extratropical variances of T is due to geostrophic structures which are explained by the variance. The unbalanced p_s variance is 1hPa in the tropics, and between 1.5 and 1.8hPa in the extratropics. This means that, although the unbalanced (T,p_s) are what counts in the tropical structure of J_b, their calibration using information averaged from all over the globe, not just from the tropical regions.

Distribution of the correlations

The vertical structure functions implied by the J_b formulation are, by construction, globally uniform for ζ and q, and almost uniform for η. The meridional distribution of the vertical (T,T) correlation with level 18 (500hPa) is shown in fig.27 (using the f-plane assumption for each latitude) and compared with the actual distribution diagnosed using the NMC method (applied for a set of narrow latitude bands). We can see that, although the correlation model is far from being perfect, there is an enormous improvement over having a uniform (T,p_s) structure as in the previous J_b formulation. The latitudinal dependence of J_b is constrained to follow a linear function of the sine of latitude, as explained earlier, so that it is impossible to reproduce exactly the real correlation distribution. Perhaps the polar analysis could be improved by refining the (T,p_s) correlations there.

The inspection of the latitudinal dependencies for the other parameters reveals that the

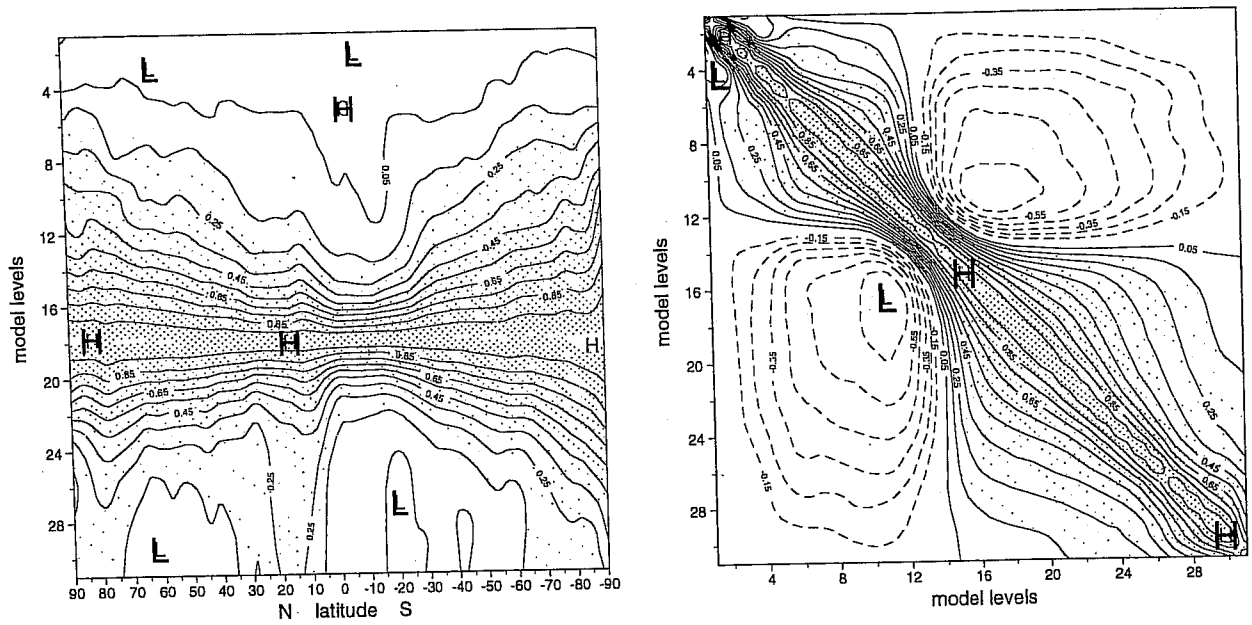


Figure 28: North-South variation of the total ζ vertical correlations at model level 18 in the NMC statistics (left panel), and vertical (T,T) correlations averaged around the North pole (right panel).

vertical correlations of η, q and unbalanced (T,p_s) are virtually uniform, so that the J_b formulation seems correct for these fields. Averages of the correlation structures over Europe has shown that, as far as the NMC method can be trusted, the J_b structures look correct and consistent with the diagnostics made in the previous sections. However, one can suspect that the real background-error covariance structures are modified locally by the density of observations, although this cannot be detected by the NMC method.

The only notable inhomogeneities were found with vorticity, and near the North pole (North of 70N ; the South Pole was not studied because some more complicated problems are likely to be generated by the high orography and low surface temperatures in this region). This is illustrated by fig.28 : the ζ vertical structures are substantially sharper near the equator than near the poles, which is quite different from the assumed picture of globally homogeneous vorticity covariances in J_b ; the polar (T,T) correlations are quite different from the extratropical ones generated by the J_b formulation (see fig.21) . This shows that locally, the structure functions of J_b are suboptimal even with respect to time-averaged error statistics. This is bound to affect the performance of the assimilation system, although it is difficult to predict exactly how. Currently it is likely that the wind analysis is too smooth in the vertical in the tropics, and that the temperature analysis is too sharp and incorrect near the tropopause in the vicinity of the poles.

7 Meteorological validation

7.1 Single-observation experiments

The effect of the J_b design on the actual analysis is usually difficult to interpret, because the structure of the analysis increments results from a complex weighted average of the observation departures from the background, if there is more than one observation in the vicinity. The effects of the J_c term and of the initialization procedure in the incremental method make the picture even more complicated.

The simplest illustration of the J_b structure functions is provided by non-incremental analyses with no J_c term and one single simulated observation per analysis. The relationship between J_b and the analysis increments is given by $J_b(\delta \mathrm{x}) = \delta \mathrm{x}^{\mathrm{T}} \mathbf{B}^{-1} \delta \mathrm{x}$ and the analysis equation

$$x^a - x^f = \mathbf{B} H^{\mathrm{T}} (H \mathbf{B} H^{\mathrm{T}} + R)^{-1} (z - H x^f)$$

where the usual notations have been used. This shows that the analysis increment is the $\mathbf{B}$ covariance structure between the observed parameter and all model fields, times a scalar. If σ_b and σ_o are, respectively, the standard background and observation errors, the value of the increment at the observation point itself is

$$H(x^a - x^f) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} (z - H x^f)$$

which makes it simple to check the value of $\sigma_b = \sqrt{H \mathbf{B} H^{\mathrm{T}}}$ if one knows the observation error and the departure from the background.

In this section the observations are simulated over oceans, where the J_c term has very little effect on the increments ; it is mainly important near orography. The effect of high-resolution initialization of the increments has a complex effect which will not be documented in this paper ; it is mainly caused by the changes in orography, but there is also a global penalization of the structures of the mass/wind balance, which may sometimes be in contradiction with the J_b structures. The large-scale part of the initialization has been switched off recently (Simmons and Rabier 1997), thereby reducing considerably the changes to the J_b structures, particularly in the tropical regions. The handling by the assimilation of the large-scales tidal waves, of the structure functions near orography and of the diabatic balance of the atmosphere are still significant problems which require further investigations (see Järvinen, Bouttier and Simmons 1997).

A sample of the J_b structure functions compared against the previous formulation is given in fig.30 and 29 . The simulated observation is a z1000 geopotential observation with $\sigma_o = 10m$ and a 10m departure from the background. This is very similar to a surface pressure observation ; the increments are small because the assumed background errors are very small for this parameter near the equator, but the implied structures could be important for the analysis of tropical storms and the use of scatterometer data. Figure 30 compares the previous and the current J_b. The T increment is exactly one would have expected from the study of the spectrally averaged (T, p_s) correlations in the previous sections. The novelty is in the implied wind structures, as seen on the maps of divergence and wind increments (with the previous J_b there was virtually no wind increment) : the wind is predominatnly divergent, with weak and more rotational

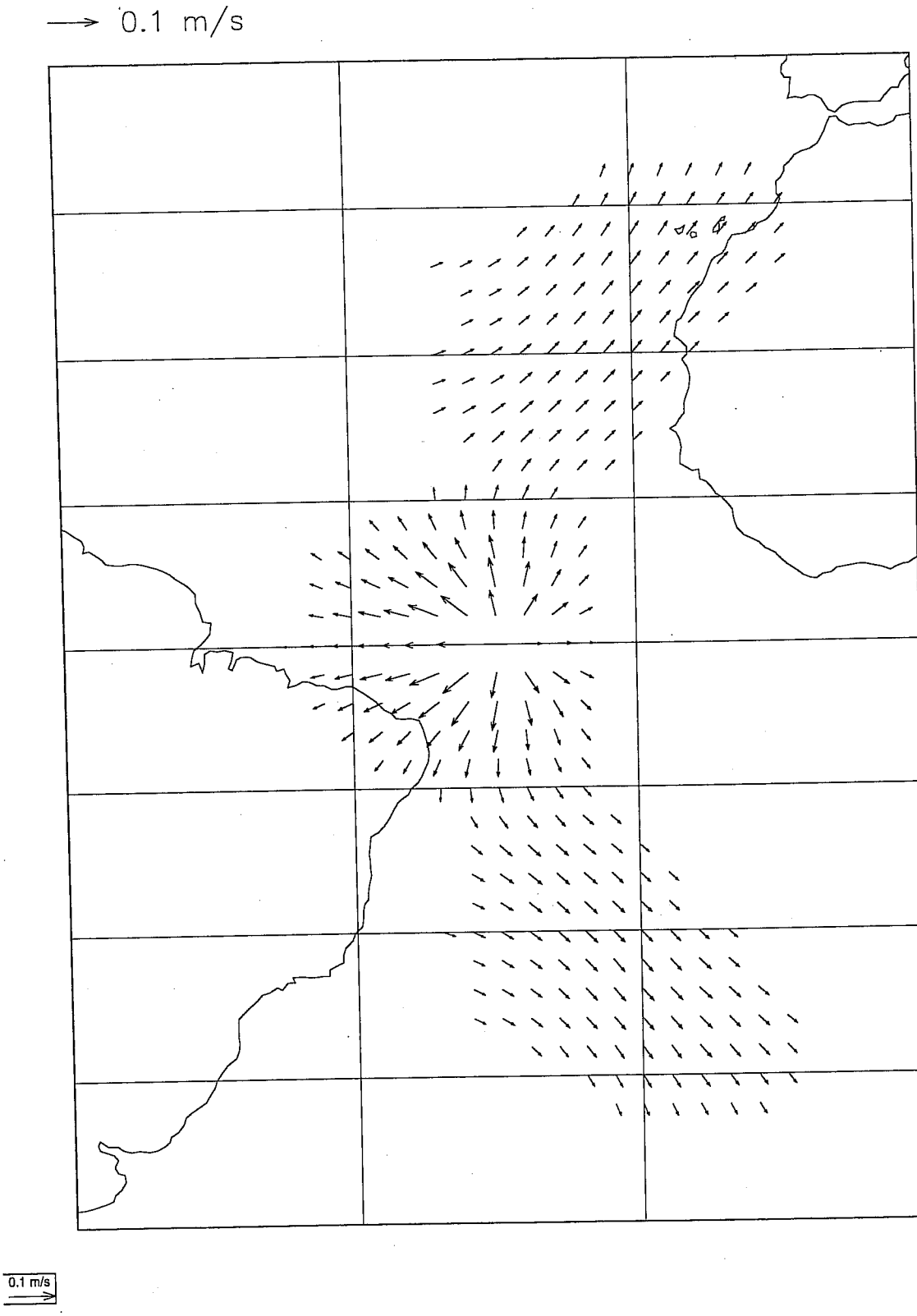


Figure 29: Wind increment generated by the simulated z1000 observation at model level 30. Some small wind increments extend far away from the observation ; they have been arbitrarily removed from the plotting.

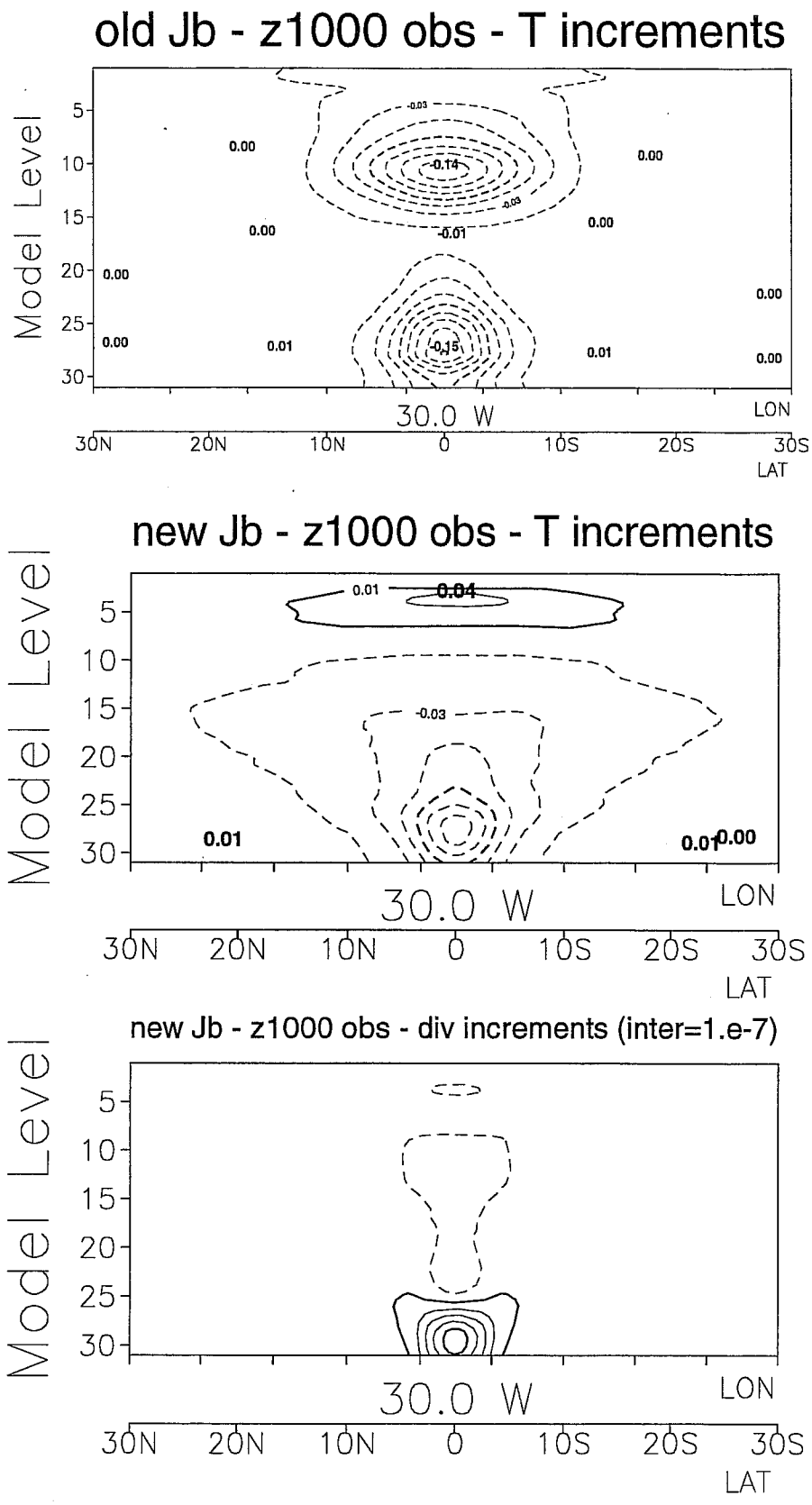


Figure 30: Vertical North-South cross-sections of increments of temperature and divergence with the previous and current ("new") J_b, generated by a simulated z1000 observation (background plus 10m), as explained by the legends.

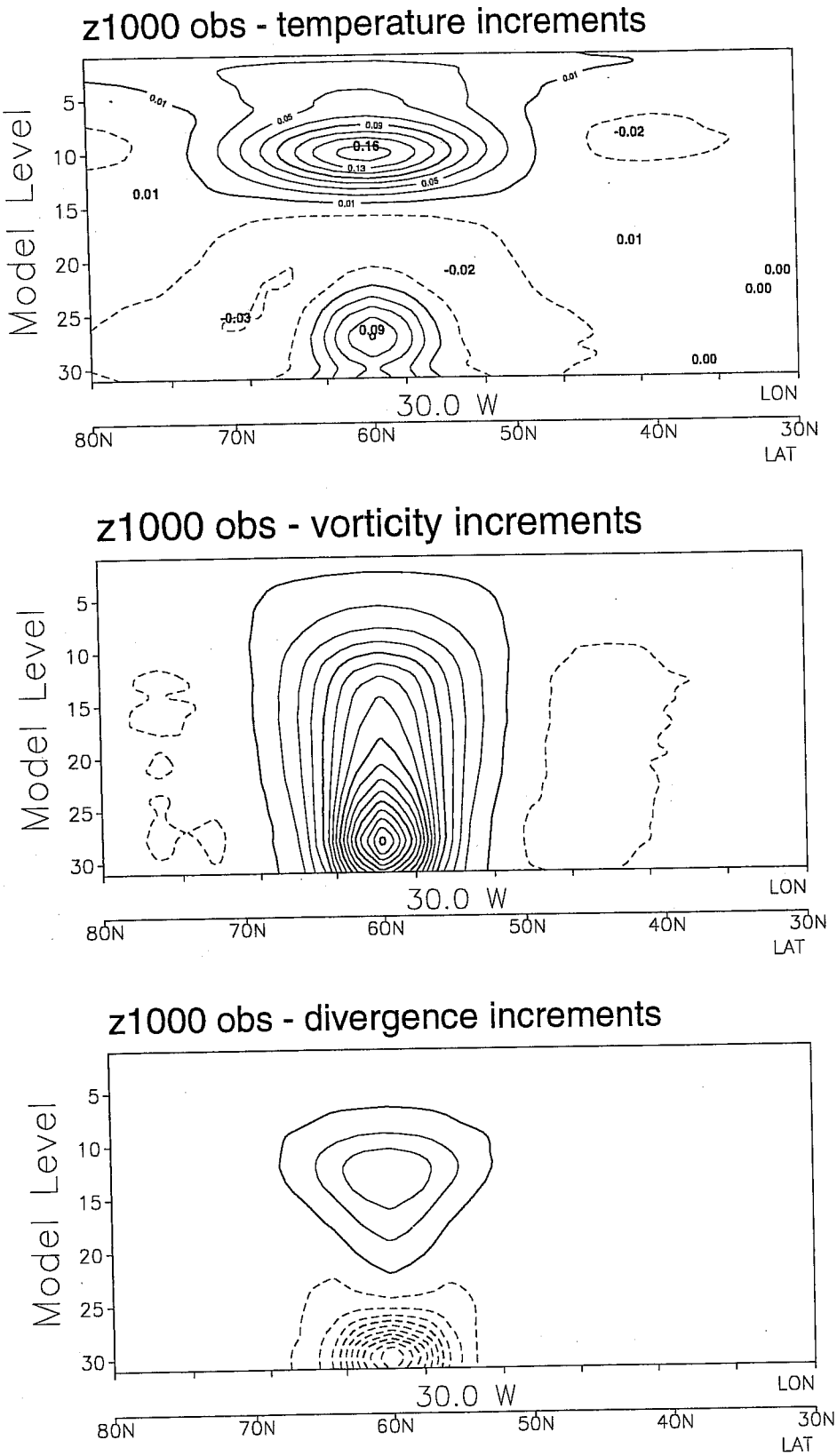


Figure 31: Temperature, vorticity and divergence increments generated by a simulated z1000 observation (10m lower than in the first guess) at 60N,30W. All increments are isotropic.

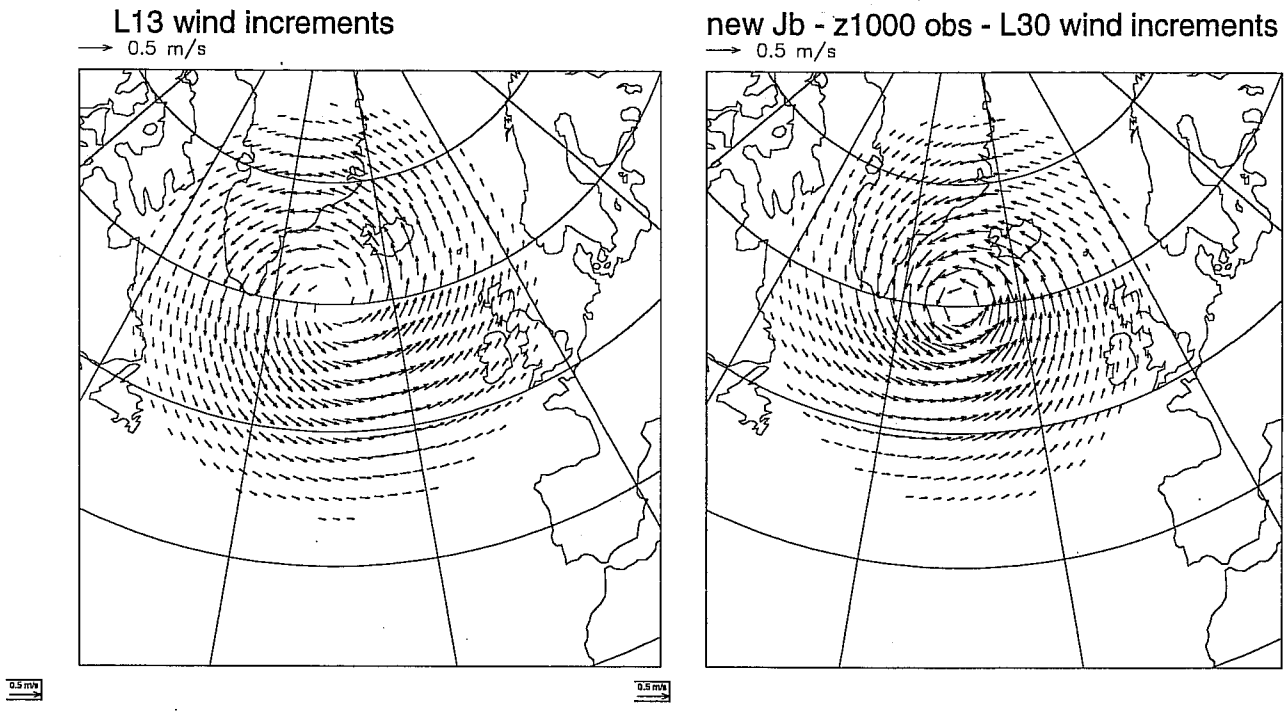


Figure 32: Wind increments at model levels 13 (300hPa) and 30 (150m) by the simulated z1000 observation at 60N,30W.

features to the North and South. The impact of activating J_c has also been tested ; it results in the addition of a negligible and complex vorticity increment.

When the z1000 observation is moved to latitude 60N (now it is 10m below the background geopotential), the increments change completely and result in the complex temperature, vorticity and divergence increments shown in fig.31 ; some wind patterns are shown in fig.32 . Again, this is completely consistent with the spectrally averaged diagnostics of correlations presented earlier. The observation results in a cyclonic increment with a warming in the lower troposphere and near the tropopause, a large barotropic vorticity increment with convergent wind near the ground and divergent wind near the tropopause. The previous J_b would have generated similar vorticity and temperature increments, but no divergence.

Figure 33 shows the increments produced by a T200 observation (similar to an AIREP), with 1K departure and observation error. There is a small divergence increment, and its combination with the vorticity structure results in the wind increment in fig.34 . Once again, this is consistent with the study of the spectral operators. The increment is comparatively less divergent than with the previous z1000 observation, because J_b assumes more geostrophic balance aloft than near the ground. The comparison with the previous J_b reveals that the wind increments were non-divergent and weaker than with the present formulation.

The increments generated by wind observations have the structures shown in fig.6 for the wind, and those implied by the resulting vorticity structures for the other parameters.

The conclusion is that the J_b structures are completely consistent with the analysis made in the section on the spectral operators : the spectrally averaged covariances are good approximations of the effective synoptic structures of the increments. The reader is referred to the relevant sections for a documentation of the complete correlation matrices, in which the

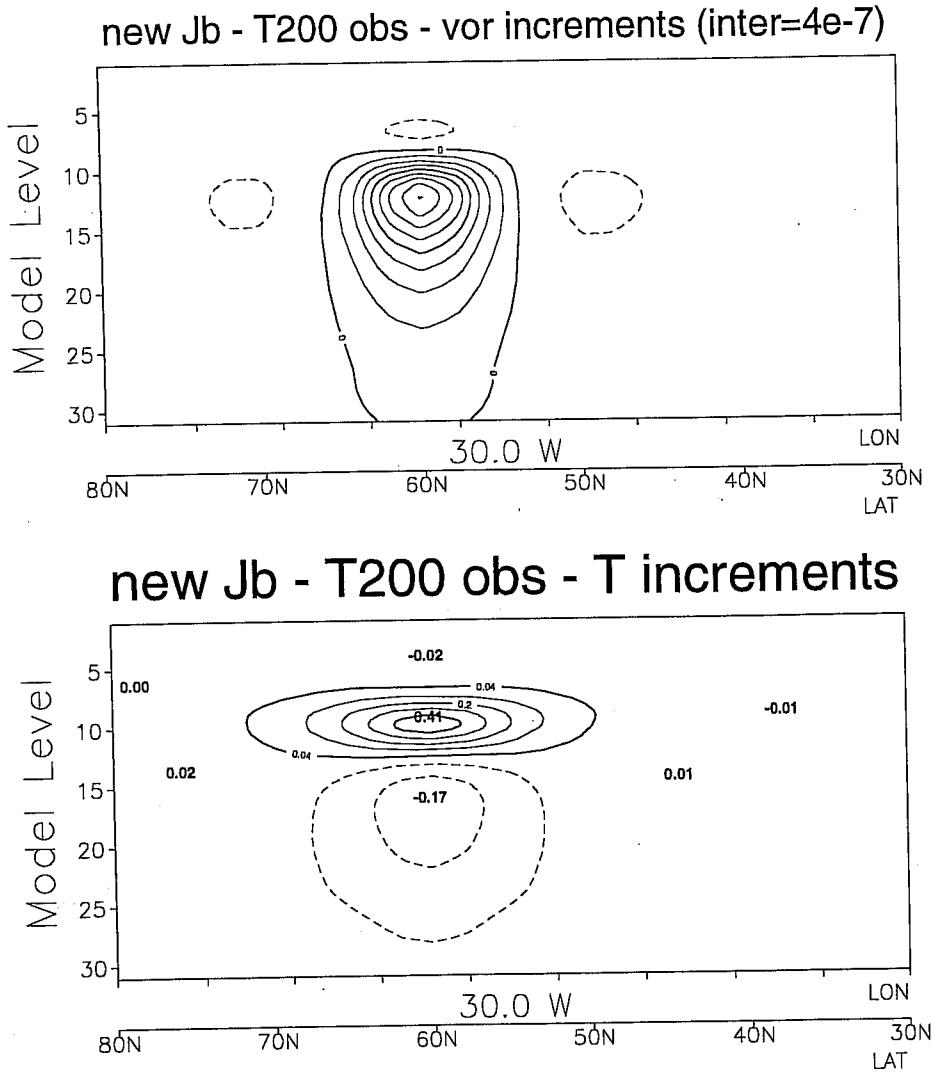


Figure 33: Temperature and vorticity increments generated by a simulated T200 observation, as explained in the legends. All increments are isotropic.

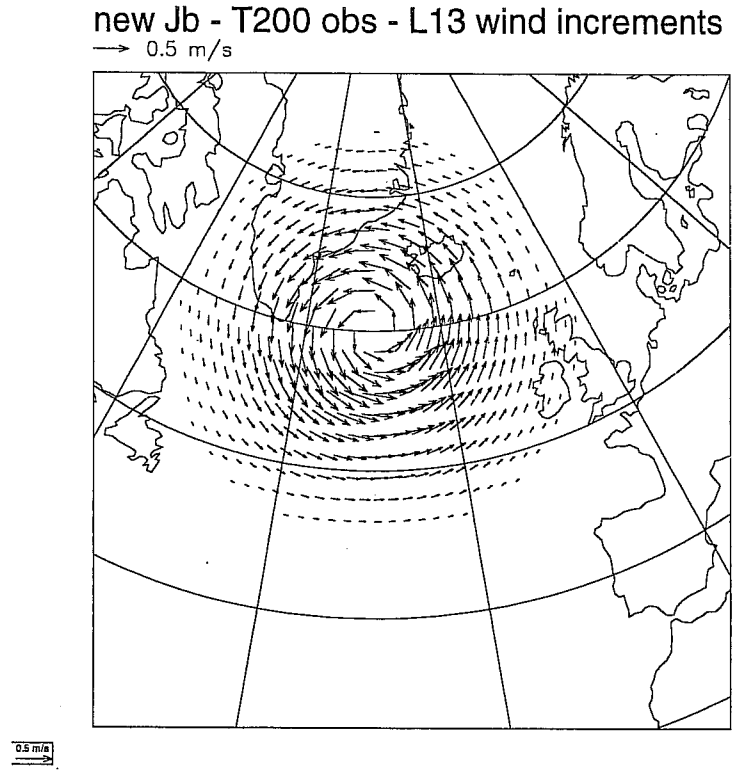


Figure 34: Wind increment (at level 13, i.e. 300hPa) generated by the simulated T200 observation.

structure functions for observations of each parameter and at any level can be seen.

7.2 Impact on the assimilation

A total of 9 weeks of experimental assimilation have been run during four seasons, plus 4 weeks of preoperational assimilation before the operational implementation of the J_b revision. This has allowed a clean comparison with the previous J_b formulation.

Although they look similar in the midlatitudes,the increments are very different from those implied by the assimilation with the previous J_b. At the start of each assimilation, the increments can be very different in a complicated way, with sometimes opposite signs, because the balance and the vertical averaging properties of the new J_b are very different. This results in a different combination of the various pieces of information provided by a localized observation of several parameters (like aircraft reports of wind and temperature) or a vertical profile of one parameter (like a PILOT sounding). After a couple of assimilation cycles, the background fields are usually different (particularly in the tropics) and a detailed comparison of the two systems is impossible.

The average amplitude of the increments is generally similar between the two systems, except in the tropical regions where it is substantially smaller with the new J_b. This is the consequence of a better average agreement between the background and the observations, however the new J_b can generate locally larger increments because of the more complex coupling between the variables (e.g. a tropical observation of mass can now generate wind increments). The better quality of the short-range forecasts is particularly apparent in the better r.m.s. fit of the tropical

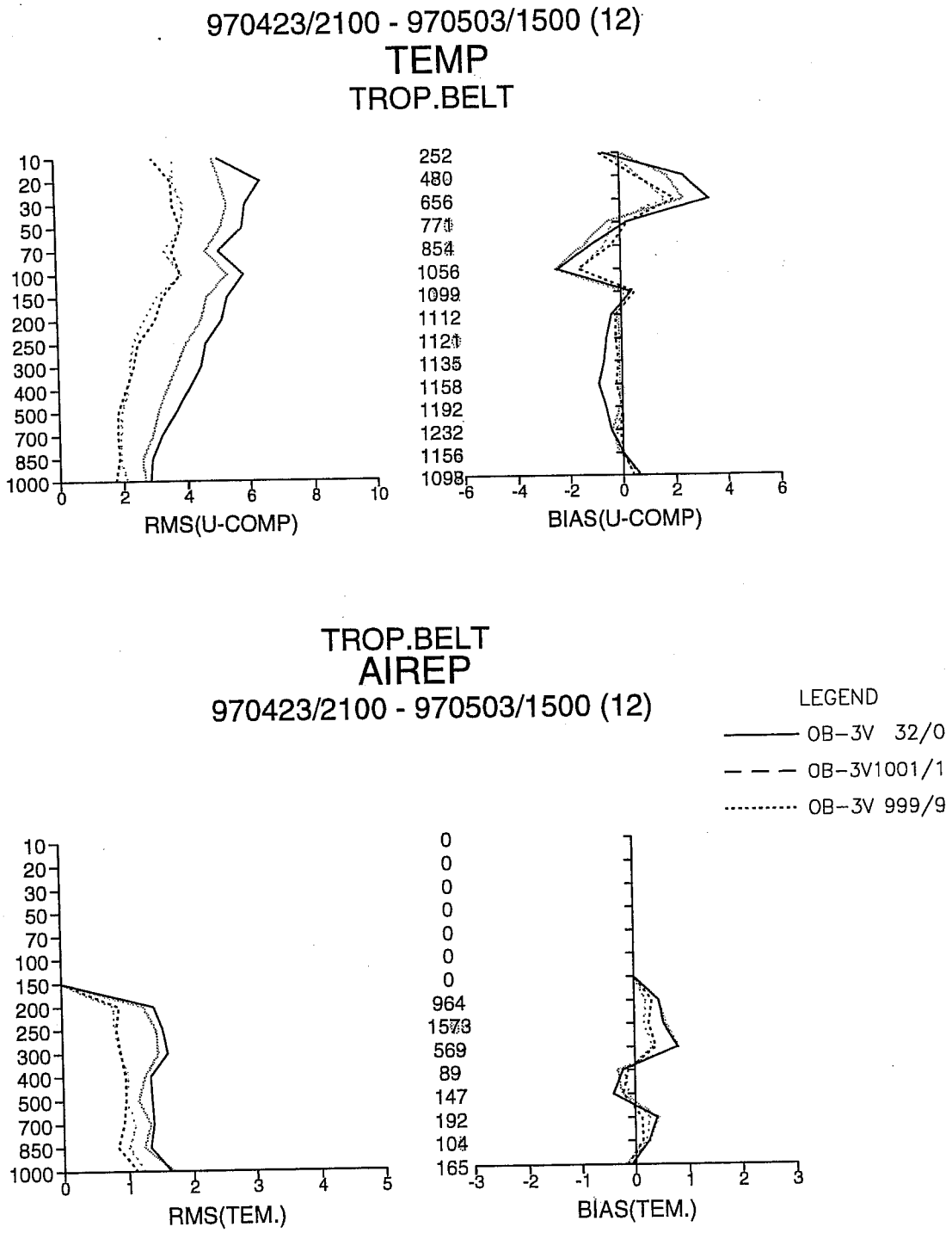


Figure 35: Impact of the J_b revision on the observation fits to the analysis (dotted curves) and background (solid curves) for the tropical TEMP winds (top panel) and the tropical AIREP temperatures (bottom panel). The black curves are for the previous operations, the grey curves are for the parallel suite. The statistics have been computed using 10 days of feedback data.

observations with the background, as shown in fig.35 ; it is a spectacular illustration that a better fit of the background temperature can be achieved by having *less* analysis fit to the observations, because the structure functions are much improved.

The "flip-flop" problem (alternance of increments with similar shapes and opposite signs in the assimilation which has a 6-hour period) is only slightly reduced by the change in the J_b formulation ; its tropical component, which is thought to be caused by discrepancies between the model tides and the observed ones, is practically unchanged as shown in a Fourier analysis of the sequences of observed surface pressure departures (Järvinen et al 1997). These problems have received a separate treatment by the suppression of the large-scale part of the initialization (Simmons and Rabier 1997) ; this change has been implemented at the same time as the new J_b formulation.

The cost of the revised J_b is less than the previous one. However, the impact on the overall cost of the 3D- or 4D-Var assimilation is small, because it is dominated by the handling of the observations (for 3D-Var) and by the direct and adjoint model runs (in 4D-Var). The improved preconditioning speeds up the experimental analyses with a small number of observations, but the convergence of the full 3D-Var is not clearly faster than with the previous J_b. Slower convergence may be a side-effect of using sharper temperature correlations in the tropics, because it increases the number of effective degrees of freedom in the variational analysis problem. There is no impact on the cost of the 4D-Var minimization because the criterion for convergence is never reached before the maximum authorized number of iterations ; in 3D-Var with the previous J_b, the convergence used to be reached a couple of iterations before the limit, now the limit of 70 iterations is always reached, with a satisfactory decrease in the norm of the gradient.

The impact on the humidity analysis and assimilation of satellite radiances is still under investigation. No obvious problem was found, and the fits to the background of both TOVS radiances and tropical humidity observations are a bit improved, although it may be just a side-effect of the overall improvement in the assimilation of wind and temperature. In the assimilations based on the old and new J_b formulations the zonally averaged temperature, humidity and Hadley cell circulations are very similar and no significant difference could be found. In the forecasts it seems that the short-range spin-up of the precipitations is slightly reduced, and that the tropospheric humidity increments are smaller in the experimental SSM/I assimilation, although all these results are very preliminary (a specific study of the assimilation of SSM/I data with the new J_b formulation will be presented in a forthcoming technical memorandum). Some work is under way at ECMWF to revise the use of TOVS radiances and it seems that the new J_b formulation behaves much better than the previous one in the stratospheric assimilation of radiances (currently SATEM retrievals from NESDIS are used in the stratosphere).

The impact of the J_b revision on the 6-hour experimental 4D-Var is very positive ; it is as large as in 3D-Var, and some problems of tropical humidity biases with the previous J_b now seem to have almost disappeared.

7.3 Impact on the forecasts

The J_b revision has a major positive impact on the quality of the forecast at all ranges. Figure 36 shows the forecast scores computed over the 3D-Var experiments that were run with the final version of the revised J_b formulation (the other experiments exhibited a positive impact

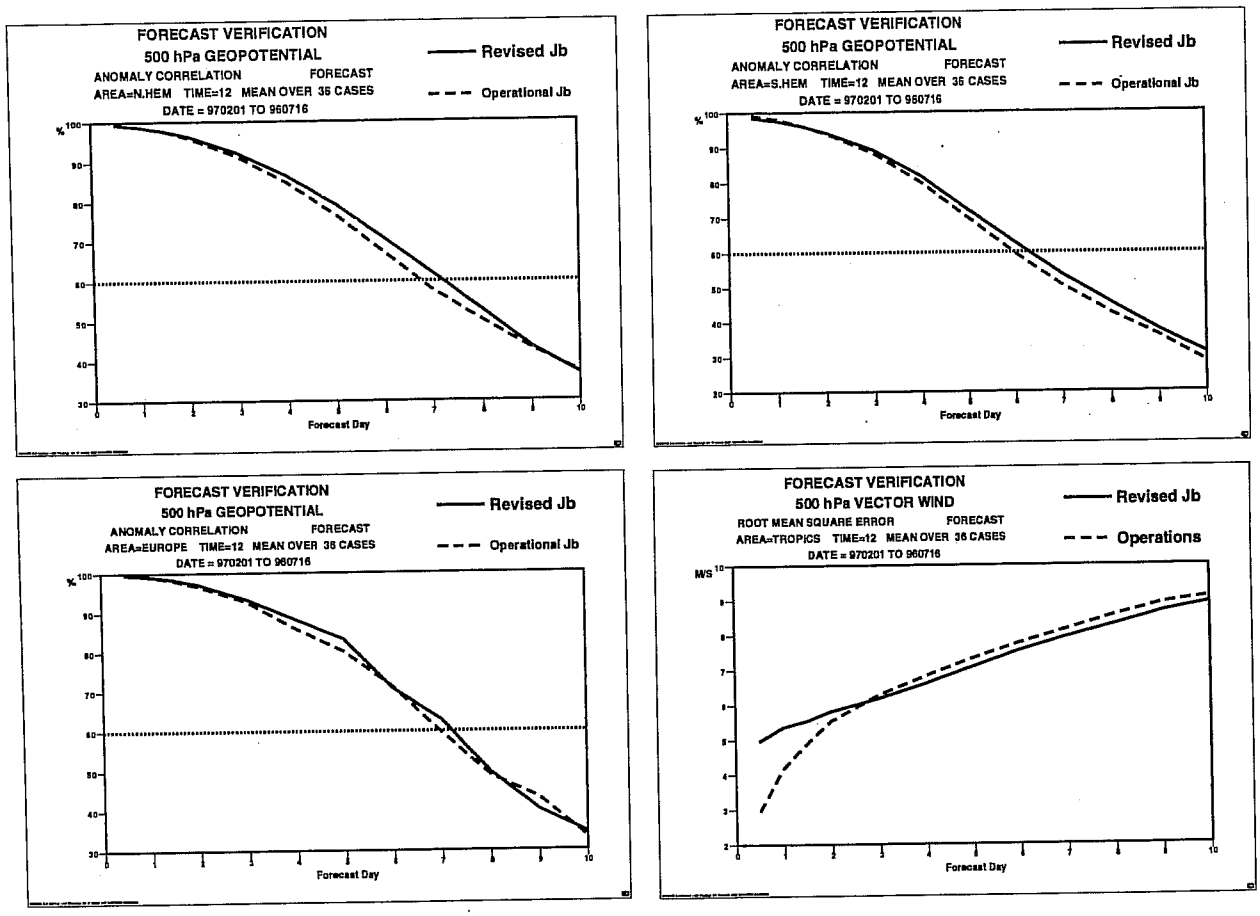


Figure 36: Impact on the forecast scores on the J_b revision (only) over 5 weeks of experimental assimilation (2 in February 1997, 3 in June/July 1996).

as well, but they were done with a less refined version of J_b). The verification is the operational analysis, which explains why the short-range tropical scores seem to have worsened : in the tropics the analyses are much changed by the J_b revision. The forecast quality is actually improved at all ranges. The medium-range forecast quality is extended by about 8 hours.

A preoperational parallel suite was run from 12 April 1997 to mid-May. Its main feature was the change in the J_b formulation, but it also included a suppression of the normal-mode initialization of the large-scale part of the high-resolution increments (Simmons and Rabier 1997), a revision to the evolution algorithm for the forecast standard errors, a change to the usage, bias correction and quality control of TOVS radiances, and minor model changes. The net impact of these modifications was very positive. All scores were significantly improved, notably the winds (at all ranges when verified against own analysis), and on all areas except the Southern Hemisphere which showed a slight degradation ; given the spread of the individual forecast scores, this is not considered to be significant. The scores over Europe are improved a lot. A sample of the scores is shown in figure 37. Several months of parallel assimilation would be required in order to assess reliably the exact impact on each individual score, but so far the bottom line is that, by all aspects, the tropical analyses and forecasts are consistently improved by the J_b revision, and that there is a small but significant positive impact on all the other scores.

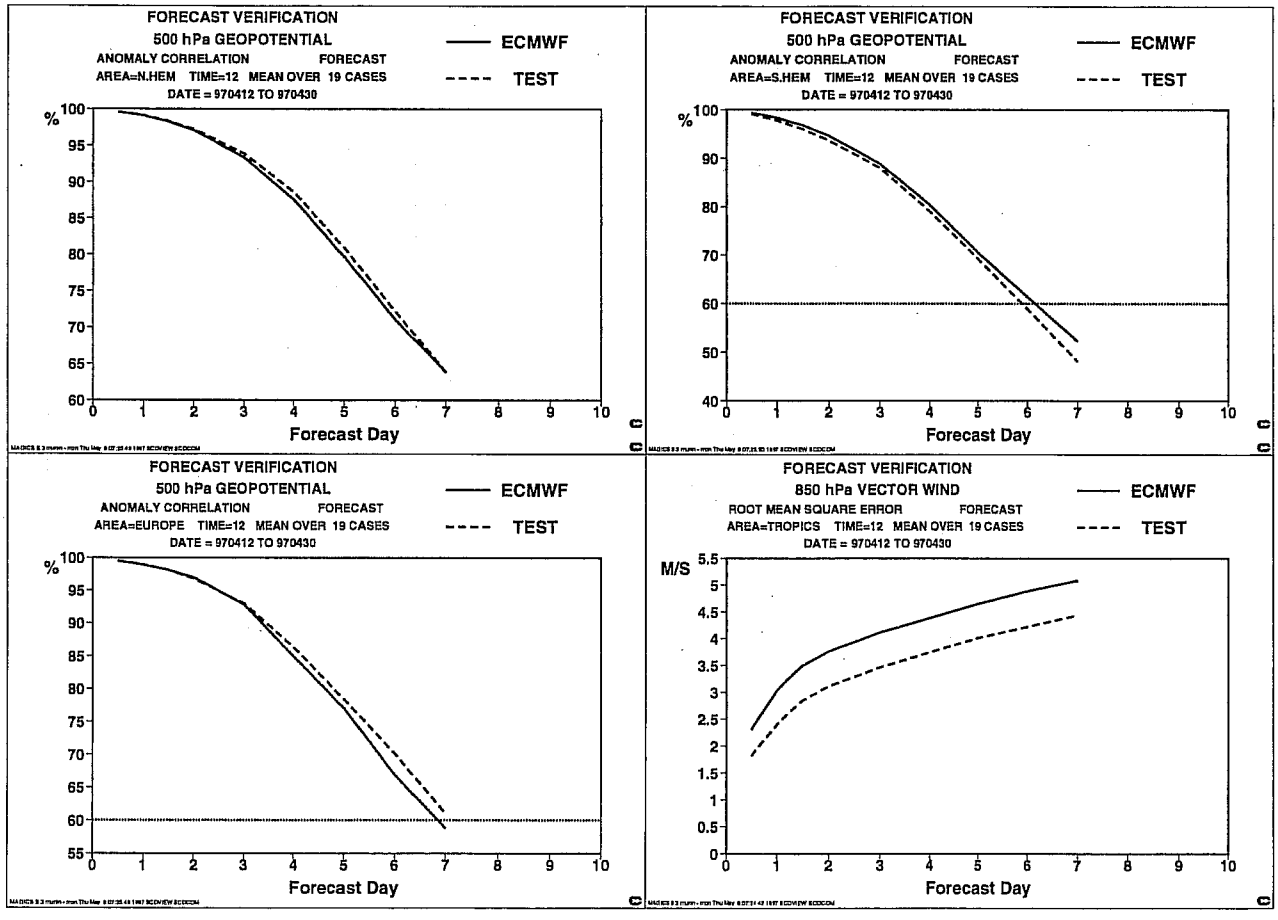


Figure 37: Impact on the forecast scores of the operational change (IFS cycle 16r2) that includes the J_b revision, over 19 days of parallel testing. The scores are computed against the own analysis of each assimilation system : the previous operational suite (solid lines), and the new suite (dashed lines).

8 Conclusion

The reformulation of J_b brings a notable improvement to the quality of the assimilation and of the ensuing forecasts. However, some weaknesses remain and need to be worked on. The structure functions are globally uniform despite the tropical/extratropical contrast which is implied by the formulation. This is clearly wrong for the vorticity field, and perhaps for the unbalanced temperature and surface pressure as well. It would be desirable to design more algebraically general balance and covariance operators in order to achieve a specific treatment of the poles, the tropics and the midlatitudes. Although it seems easy to do in the balance operators (it is possible to combine several versions of them through geographical masks), enforcing a geographical variability of the covariances raises some complicated mathematical issues.

From a theoretical point of view, the J_b reformulation is based on the hypothesis that the balanced component of the short-range forecast errors can be expressed as a function of vorticity, the vorticity field itself being assumed to be always balanced. This is not equivalent to the concept of a balanced variety defined by the Rossby-gravity normal modes of the model. Some authors have advocated the use of the Hough space representation, which might provide better results : it would be interesting to try. The algebraic constraints on the operators could also be relaxed a bit, as mentioned above in order to allow more geographical structure, and also a different kind of mass/wind balance in the tropics, or a different treatment of the very large scales.

The analysis of humidity has not been modified. This is clearly a major weakness of the current assimilation system. A better humidity variable (like the logarithm of q) should be used, and a balance between humidity and the other variables would be useful ; it probably requires a distinct treatment of the tropics and the extratropics. Finally, the dynamical initialization procedure should be diabatic ; it seems that digital filtering could be more appropriate than normal mode initialization, following the experience at other numerical prediction centres.

Some work is being done at ECMWF on the assimilation of ozone. Preliminary experiments have shown that a significant part of the ozone field can be coupled multivariately with the wind analysis. It means that in the near future, the ozone analysis will benefit from wind observations, and the stratospheric wind analysis might be improved thanks to the introduction of ozone measurements.

The success of the assimilation of satellite data, e.g. TOVS radiances, depends on the quality of the J_b constraint. It seems that the revised J_b will allow a better use of TOVS data, and hopefully some kind of bias correction scheme will be integrated within the atmospheric J_b term of the variational analysis.

The basic characteristics of J_b currently depend a lot on assumptions made about the NMC method. It is necessary to monitor their correctness using independent statistics of observation departures e.g. to monitor the amplitude of background error variances of correlation length-scales in data-rich areas.

The improvements in J_b benefit 4D-Var and the Kalman filter directly. Technically the knowledge of the exact square root of the background error covariance matrix facilitates the work on numerical preconditioning and simplifications to the Kalman filter. However, there are serious doubts as to how appropriate the J_b formulation is for the generation of useful flow-

dependent structure functions, because it is too barotropic. In dynamically unstable areas, we need to make the J_b covariances more isotropic in phase space (i.e. with sharper correlations in physical space).

All these important problems are being investigated at ECMWF.

References

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljević, D., Branković, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N. and Viterbo, P, 1997: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. Submitted to *Quart. J. Roy. Met. Soc.*

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, A. Hollingsworth, M. Fisher, F. Rabier, 1997: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part 1: formulation. Submitted to *Quart. J. Roy. Met. Soc.*

Daley, R., 1997: Generation of global multivariate error covariances by singular-value decomposition of the linear balance equation. *Mon. Wea. Rev.*, **124**, 2574-2587.

Fisher, M. and P. Courtier, 1995: Estimating the covariance matrix of analysis and forecast error in variational data assimilation. *ECMWF Res. Dept. Tech. Memo. no.220,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Heckley, W., P. Courtier, J. Pailleux and E. Andersson, 1992: The ECMWF variational analysis: general formulation and use of background information. *Proceeding of the ECMWF workshop on variational assimilation, 9-12 November 1992, pp.49-93.* Available from ECMWF, Shinfield Park, Reading RG2 9AX, G.-B.

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.

Järvinen, H., F. Bouttier and A. Simmons, 1997 : Four-dimensional variational assimilation of tropical atmosphere: autocorrelation properties and optimality. *Submitted to Quart. J. Roy. Met. Soc.*, 13pp.

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

Rabier, F., and T. Mc Nally, 1993 : Evaluation of Forecast Error Covariance Matrix. *ECMWF Res. Dept. Tech. Memo. no.195,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth and F. Bouttier, 1997: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part II: structure functions. *Submitted to Quart. J. Roy. Met. Soc.*

Simmons, A. and F. Rabier, 1997: Removal of the incremental initialization of medium and large scales in the ECMWF analysis. *ECMWF RD internal memorandum R43/AJS/51/DA,* 22 May 1997, 12pp.