INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR/534 19

ICTP/WMO WORKSHOP ON EXTRA-TROPICAL AND TROPICAL
LIMITED AREA MODELLING
22 October - 3 November1990

"General Problems of Data Assimilation: II"

P. UNDEN
ECMRWF
Reading, U.K.

METHODS OF DATA ASSIMILATION

By P.Unden, European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, England.

1. INTRODUCTION

1.1 Summary

These lectures are intended to give insight into the methods used in metecrological data analysis. The aims of the analysis will be discussed and what tools are available to make the best possible analysis. Different methods are reviewed although most of the in-depth discussion will be concerned with optimum interpolation since this is the most widely used method. The problem making use of the time dimension is also discussed and the adjoint method of assimilation is a recent interesting development in this area. Several parts of these note make use of a lecture note by Lonnberg and Hollingsworth (1984)

1.2 Aims

A numerical analysis of the state of the atmosphere is needed for the following purposes:

1) Initialising a numerical forecast. In this case the best analysis is the one which gives the best numerical forecast. Hence the analysis scheme may need to be tuned for the particular forecast model which it is used with and used for.

2) Diagnostic studies of the atmosphere, displays in the forms of plotted charts, archives and forecast verification. The best analysis in this context may be different from 1). Systematic biases need to be avoided for diagnostic and climatological studies. When a forecast model is used for the first guess this influences the analysis a lot and has to be born in mind.

3) Observation checking. For all practical schemes this is a very important problem. Many of the observations with large errors may be rejected by the other quality control methods mentioned when discussing observations but the best final check is against an analysis made without the datum to be checked.

1.3 Data available

1) Observations. These have been discussed in a previous lecture as well as in the lecture note about observations.

2) First guess. Most analysis methods use a preliminary estimate of the field being analysed which is changed during the analysis procedure only at and around observations. Usually this is done by analysing departures from the first guess field rather than the total values. The possible fields which can be used as first guess are:

a) climatology, i.e. the average analysis for the appropriate season

b) persistence, i.e. previous analysis

c) forecast, i.e. a numerical forecast from the previous analysis

For a particular analysis one or a combination of these fields may be chosen depending on the application.

3) Knowledge of the likely structures and scales of atmospheric motion. This information is often incorporated inplicitly into analysis schemes without being clearly stated or quantified. Such knowledge is often expressed as relationships which the atmosphere obeys to certain approximations and these may be used as constraints on the analysis. The following list of conditions have been used in analysis schemes:

a) Increments to atmospheric fields are smooth and continuous. This is done in spite of the existence of fronts.

b) The value being analysed is known to be of certain scales. Thus a single observation will influence the analysis on the most likely scales or scale and dense inaccurate observations will be averaged over these scales in an ideal analysis scheme.

c) The atmosphere is in hydrostatic balance.

d) The is in geostrophic or gradient wind balance.

e) The horizontal wind is non-divergent.

f) The atmosphere is in a state which satisfies the balance equation.

g) The atmosphere is convectively stable.

g) The atmosphere is not super-saturated.

There are, however, other well known features of the atmosphere which are much more difficult to make use of in an analysis scheme and therefore not yet used in any practical schemes:

1) Baroclinic developing systems have a certain tilt with height.

2) Warm sectors in baroclinic systems have a certain shape.

2. CONTINUOUS AND INTERMITTENT ASSIMILATION

One can make use of a previous analysis or a forecast from a previous analysis to introduce a time dimension into the analysis. This is really the meaning of data assimilation. Such a first guess may give a good estimate of the current state of the atmosphere and observations are used to make small adjustments to the first guess field. As the procedure is repeated for subsequent observation times the analyses should improve more and more as more data are influencing the state of the atmosphere through the "memory" of the first guesses. Normally the first guess is constructed by using a forecast model since most of the time evolution during say a 6-hour period can be explained by the model. In this context there are basically two different ways of doing data assimilation. Many observations are made at so called synoptic times every 3 or 6 hours for (SYNOP) or mainly every 12 hours for TEMP's and PILOT's. Other observations are avaiable at irregular times like satellite and aircraft data. For both cases one should ideally introduce observations and analyse them at using the first guess at the appropriate time and continue the forecast from the updated state of the atmosphere at each observation time. This is called continuous assimilation.

There are basically two problems with this approach. Splitting the observations from a 6-hour period into sets for each timestep often results in very few data being available in the analysis simultaneously and one cannot exploit data redundancy and interrelations fully. Also the analysed state of the atmosphere needs to be initialised for the model to have a balanced state and this is expensive if performed repeatedly.

The intermittent method of data assimilation gathers all observations during a time period and treats them as being valid for the same time (except maybe some corrections of observed values when possible) and analyses the atmosphere using all those data simulaneously. The relationships between wind and pressure/height data and between surface and upper air data can then be exploited.

These methods are both often referred to as 4-dimensional analysis but observations in the time dimension do not enter the analysis in the same way as those for the space dimensions. Their effect are only carried forward by the forecast model in an indirect way and it would be more appropriate to call it a 3 1/2-dimensional analysis.

3. THE SUCCESSIVE CORRECTION METHOD

The succesive correction method was the first automated analysis method as developed by Berthorsson and Doos (1955) and is also often referred to as Cressman analysis after Cressman (1959).

The first guess is constructed from a forecast or a combination of forecast, previous analysis and climatology depending of the perceived accuracy of the different fields. Departures between observations, A_i^o, and the first guess values, A_i^p, are formed and the analysed departure at a first guess grid point, A_k^i, is then a linear combination of all the observations.

$$A_k^i - A_k^p = \sum_{i=1}^{n} w_{ki} (A_i^o - A_i^p)$$

The weights w_{ki} of the observations are determined by a function of distance between observation and gridpoint, c_{ki}. They are then normalised by the sum of the weights for all observations and optionally a weight for the first guess, c^p. Without this weight the first guess would not be used at all.

$$w_{ki} = \frac{c_{ki}}{c^p + \sum_i^k c_{ki}}$$

In practice a further correction is neccessary to allow for unever distribution of data.

The distance function falls close to zero at some radius and only observations within the circle need to be considered for any one gridpoint. The process can be repeated with reduced such radie in order to analyse finer scales in repeated scans.

The method has been quite successive especially where enough data are available to define the atmosphere and is very cheap in computation and easily programmed. It is however very empirical in its nature and proper use of the known properties of the atmosphere cannot be made.

4. VARIATIONAL ANALYSIS USING THE ADJOINT METHOD

Variational techniques have been suggested for use in data assimilation in various contexts over the last decades. In operational systems such methods have mainly been used as adjust increments of different variables analysed separately. Renewed interest has emerged since the recent developments extending the variational method in the time dimension by using the adjoint of a forecast model. The desribtion of the method here follows Dimet and Talagrand (1986).

The variational problem is to find a state Y(t) which minimizes the functional

$$J[Y(t)] = \sum_{i=1}^{n} < Y(t_i) - \hat{Y}(t_i) , Y(t_i) - \hat{Y}(t_i) > \qquad (4.1)$$

where $< , >$ denotes the scalar product defined for the space we are considering. The observations $\hat{Y}(t_i)$ are available at times $t_1, t_2, \ldots, t_n$ and we seek the solution which minimizes J over the whole time period t_1, t_n. The time evolution of Y is desribed by a forecast model H:

$$\frac{dY}{dt} = H(Y) \qquad (4.2)$$

The values $Y(t_i)$ in (1) are thus computed using (4.2). The minimum of J will not in general be zero since the observations cannot be expected to be exactly compatible with the state described by the forecast model.

The problem formulated by (4.1) and (4.2) is uniquely defined by by the state vector Y(t) for any t in the time interval considered. Then it is sufficient to determine the minimum of J at one time step, say the initial time and find an $Y(t_1)$ which minmizes the cost function J over the whole time interval $t_1 - t_n$. The first order variation of J with respect to a variation of $Y(t_i)$ is

$$\delta J = 2 \sum_{i=1}^{n} < Y(t_i) - \hat{Y}(t_i) , \delta Y(t_i) > \qquad (4.3)$$

The variations $Y(t_i)$ are obtained by integrating the tangent linear equation

$$\frac{d\delta Y}{dt} = A(t)\, \delta Y \qquad (4.4)$$

starting from some initial variation $\delta Y(t_1)$. This equation is obtained by linearizing the forecast model around the trajectory Y(t) for each time step. The coefficients of A are made up of the values of the derivatives of H with respect to Y. The actual values of Y from the full system integrated by (4.2) from a first guess state are thus used at each time step for defining the derivatives. The equation (4.4) describes the evolution along a tangent plane to the original model at each time step. The time evolution of (4.4) will not be the same as integrating (4.2) with the initial condition $Y(t_1)+\delta Y(t_1)$ but is a good aproximation to it provided the departue $\delta Y(t_1)$ is small. The coefficients in A are fixed and independent of the initial variation

$\delta Y(t_1)$. $\delta Y(t)$ at any time step is thus linearly dependent on the initial condition and this linear operator is called the resolvent $R(t_i, t_1)$ between times t_i and t_1.

$$\delta Y(t_i) = R(t_i, t_1)\, \delta Y(t_1) \qquad (4.5)$$

Substituting into (4.3) gives

$$\delta J = 2 \sum_{i=1}^{n} \langle Y(t_i) - \hat{Y}(t_i),\ R(t_i, t)\, \delta Y(t_1) \rangle \qquad (4.6)$$

Now, the second term in (4.6) involves the resolvent of the variation for all the time steps involved but we would like to know the variation $\delta Y(t_1)$ itself which minimizes J in (4.3). A powerful tool in linear algebra to achieve this is to use the adjoint of the linear operator R. The adjoint of a linear operator L is defined by the following equality using the scalar product for any two vectors X and Z in the space considered.

$$\langle \underline{X}, L\underline{Z} \rangle = \langle L^* \underline{X}, \underline{Z} \rangle \qquad (4.7)$$

Using this relationship and introducing R^* as the adjoint of R gives

$$\delta J = \langle 2 \sum_{i=1}^{n} R^*(t_i, t_1) [Y(t_i) - \hat{Y}(t_i)],\ \delta Y(t_1) \rangle \qquad (4.8)$$

This is acually a definition of the gradient of J. The scalar product between the gradient and the variation of the independent variable (Y) gives the variation of the function (J). The gradient of J is thus

$$\nabla J = 2 \sum_{i=1}^{n} R^*(t_i, t_1) [Y(t_i) - \hat{Y}(t_i)] \qquad (4.9)$$

In order to find R^* the adjoint tangent linear equation is introduced here:

$$\frac{d\, \delta Y^*}{dt} = -A^*(t)\, \delta^* Y \qquad (4.10)$$

δY^* is a different variation of Y and A^* is the adjoint of A. The resolvent of (4.10) between times t' and t is called $S(t, t')$. For any two solutions of the direct and adjoint tangent linear equations, (4.4) and (4.10) respectively, their scalar product is constant with time since

$$\frac{d}{dt} \langle X(t), Z(t) \rangle = \langle \frac{dX}{dt}, Z(t) \rangle + \langle X(t), \frac{dZ}{dt} \rangle =$$

$$= \langle A(t) X(t), Z(t) \rangle + \langle X(t), -A^*(t) Z(t) \rangle = 0 \qquad (4.11)$$

This follows directly from the definition of the adjoint (4.7) and the identity can be used to compare a solution of the direct tangent linear equation (4.4) at time t_i from the initial condition $X(t_1)$ with the solution of the adjoint equation (4.10) with initial condition $Z(t_i)$. The scalar products between the two states at t_1 and at t_i must be the same.

$$\langle X(t_1),\ S(t_1, t_i) Z(t_i) \rangle = \langle R(t_i, t_1) X(t_1),\ Z(t_i) \rangle \qquad (4.12)$$

It then follows from the definition of adjoint that the desired R^* in (4.8) is actually the resolvent of the adjoint equation S.

$$S(t_1,t_i) = R^*(t_i,t_1) \qquad (4.13)$$

and

$$\delta y^*(t_1) = R^*(t_i,t_1)\,\delta y^*(t_i) \qquad (4.14)$$

Having found the adjoint of R the term $R^*(t_i,t_1)[Y(t_i)-\hat{Y}(t_i)]$ can be computed by a backwards integration of (4.10) from t_i to t_1. This is done for each time step and observation and contributions for time step t_i from the departures $Y(t_i)-\hat{Y}(t_i)$ resolved by R^* are accumulated. The final result is half the gradient of of J. This gradient is then used in a decent algorithm to find a lower value of J.

In practice the procedure starts by inegrating the forecast model from a first guess and storing the values of Y at each time step. These will be used for forming the coefficients in A . The adjoint resolvent can then be computed and contributions from each observation are accumulated. The decent algorithm finds a lower value of J but this is usually not the minimum. The new values of Y(t) are used for another integration of the forecast model and new departures calculated to compute a new estimate of the gradient. The iterative process continues until a satisfactory minimum has been reached.

Practical trials by Courtier and Talagrand (1987) has shown that these existing decent algorithms work well for a low resolution barotropic model. The resulting analysis is realistic provided the data coverage is good and the time interval long enough. Improvements were noticed when a noise constraint was added to the cost function J.

This method is very demanding in terms of computer resources since quite a few iterations may be necessary in a full model. The main concerns are the cost of repeated model integrations themselves, storage of all the trajectory Y(t) and the coefficients for A and the cost and storage needed for the decent algorithm. There are however many attractions which makes research in this area important for future data assimilation systems. It is the only tractable way of really including the time dimension and using the observations according to their sensitivity with respect to the initial state of the model. Information is spread both forwards and backwards in time whereas today's 3 1/2-dimensional schemes only can convey information forwards in time.

The variational approach has a lot of flexibility. The cost function can be split into many independent terms for e.g. individual observation types, model balance and other possible constraints. Also the way that the observed data are realated to the model variables can be considered correctly when calculating the gradient. An important example of this is the usage of satellite radiation data.

5. OPTIMUM INTERPOLATION

5.1 Introduction

Optimum interpolation is also sometimes called statistical interpolation since the interpolation formula depends on the knowledge on knowledge of the first and second order statistical moments of the fields involved. It is only optimal in a linear sense if the statistics is exact but this is of course never the case. The technique is usually credited to Gandin (1963)

5.2 Statistical concepts

a) Expected value. This is the average of a large number of realisations of observations or pairs of observations denoted by $\langle \quad \rangle$.

b) True value. This is the actual value of the atmospheric state after the scales we do not wish to analyse have been removed. It is thus a somewhat artificial concept but it should be thought of as the best approximation of the real state that the assimilation system can possibly achieve with its current resolution. The true value is denoted by the super-script t and observations have super-scripts o and first guess predictions use p. It is assumed that observations and first guesses are unbiased.

c) Error. Departures from the true values for observations and first guess values. The variance of such errors is the expected value of the squares of such errors and denoted by E^2. For prediction and observation errors they are:

$$E^{p2} = \langle (A^p - A^t)^2 \rangle = \langle a^{p2} \rangle$$

$$E^{o2} = \langle (A^o - A^t)^2 \rangle = \langle a^{o2} \rangle$$

d) Covariances. For two points i and j the prediction error covariance is $\langle a_i^p a_j^p \rangle$ and the observation error covariance $\langle a_i^o a_j^o \rangle$. The prediction error - observation error covariance $\langle a_i^p a_j^o \rangle$ is assumed to be zero here.

e) Correlations are obtained by dividing the covariances by the two square-roots of the involved variances (standard deviations). The correlations are thus normalised; i.e. they are always in the range [1,-1]. They are used in most practical schemes instead of covariances but they do not necessarily have to be used.

f) Homogeneity. A statistical property of a meteorological field is homogeneous if its dependence on the two positions is independent of a translation of those two positions, i.e. the property is only dependent on the separation and direction but not of the positions of

the two points.

g) Isotropy. A statistical property (e.g. variance of covariance) is isotropic if it is indepndent of direction.

5.3 Optimum interpolation formulas

The analysis is formed through a linear combination of observational departures from the first guess with some weights w_i for each observation i:

$$A_k^i - A_k^p = \sum_{i=1}^{N} w_{ki} \, (A_i^o - A_i^p) \qquad (5.3.1)$$

Subtracting the true values A from both sides and adding and subtracting the true values at the observation points inside the sum gives:

$$A_k^i - A_k^t = A_k^p - A_k^t + \sum_{i=1}^{N} w_{ki} \, (A_i^o - A_i^t - (A_i^p - A_i^t)) \qquad (5.3.2)$$

The subscript i means that the values are at the observation positions whereas the analysed or interpolated value A_k^i is generally not at an observation point. Normally the analysis is required at the model's gridpoints. It is covenient to normalize the departures by the prediction error standard deviation when we are dealing with homogenous cases.

$$\frac{A_k^i - A_k^t}{E_k^p} = \frac{A_k^p - A_k^t}{E_k^p} + \sum_{i=1}^{N} w_{ki} \left(\frac{A_i^o - A_i^t}{E_i^p} - \frac{A_i^p - A_i^t}{E_i^p} \right)$$

The normalized values are

$$\alpha_k^i = (A_k^i - A_k^t)/E_k^i$$

$$\alpha_i^p = (A_i^p - A_i^t)/E_i^p$$

$$\alpha_i^o = (A_i^o - A_i^t)/E_i^o \quad ; \quad \varepsilon_i^o = E_i^o/E_i^p \; ; \; \varepsilon_k^i = E_k^i/E^p$$

$$\alpha_k^i \varepsilon_k^i = \alpha_k^p + \sum_{i=1}^{N} w_{ki} \, (\alpha_i^o \varepsilon_i^o - \alpha_i^p) \qquad (5.3.3)$$

The interpolation error on the left hand side of (5.3.3) is then squared to give us an expression for the squared interpolation error.

$$(\alpha_k^i)^2 (\varepsilon_k^i)^2 = (\alpha_k^p)^2 + 2 \sum_{i=1}^{N} w_{ki} \left[(\alpha_k^p \alpha_i^o) \varepsilon_i^o - \alpha_k^p \alpha_i^p \right] +$$

$$+ \sum_{j=1}^{N} \Bigg\{ w_{kj} \, \alpha_j^o \varepsilon_j^o \sum_{i=1}^{N} w_{ki} \, \alpha_i^o \varepsilon_i^o - w_{kj} \, \alpha_j^p \sum_{i=1}^{N} w_{ki} \, \alpha_i^o \varepsilon_i^o -$$

$$- w_{kj} \, \alpha_j^o \varepsilon_j^o \sum_{i=1}^{N} w_{ki} \, \alpha_i^p + w_{kj} \, \alpha_j^p \sum_{i=1}^{N} w_{ki} \, \alpha_i^p \qquad (5.3.4)$$

This positive quantity should be minimized. However we can only hope to do this in a statistical sense, i.e. over a large number of cases. This is done by forming the expected values of all terms in (5.3.4). The left hand side is then the normalized interpolation error.

$$\varepsilon_k^{i\,2} = 1 + 2\sum_{i=1}^{N} w_{ki}\left(\langle \alpha_k^P \alpha_i^0\rangle \cdot \varepsilon_i^0 \cdots \langle \alpha_K^P \alpha_i^P\rangle\right) +$$

$$+ \sum_{j=1}^{N} w_{kj}\left\{\sum_{i=1}^{N} w_{ki}\left(\langle \alpha_j^0 \alpha_i^0\rangle \cdot \varepsilon_j^0 \varepsilon_i^0 - \langle \alpha_j^P \alpha_i^0\rangle \varepsilon_i^0 - \right.\right.$$

$$\left.\left. - \langle \alpha_j^0 \alpha_i^P\rangle \varepsilon_j^0 + \langle \alpha_j^P \alpha_i^P\rangle\right)\right\} \qquad (5.3.5)$$

Assuming that the prediction errors are uncorrelated with the observation errors gives:

$$\varepsilon_k^{i\,2} = 1 - 2\sum_{i=1}^{N} w_{ki}\langle \alpha_k^P \alpha_i^P\rangle + \sum_{j=1}^{N} w_{kj}\left\{\sum_{i=1}^{N} w_{ki}\left(\langle \alpha_j^P \alpha_i^P\rangle + \right.\right.$$

$$\left.\left. + \langle \alpha_j^0 \alpha_i^0\rangle \varepsilon_j^0 \varepsilon_i^0\right)\right\} \qquad (5.3.6)$$

This error should be as small as possible and the weights w can be varied for each observation so that ε becomes smaller. The minimum of ε with respect to each observational weight is found by taking the partial derivatives with respect to each weight w_{kj} and setting that to zero. This leads to N equations, one for each observation. The one for observation j is:

$$0 = -2\langle \alpha_k^P \alpha_j^P\rangle + 2\sum_{i=1}^{N} w_{ki}\left(\langle \alpha_j^P \alpha_i^P\rangle + \right.$$

$$\left. + \langle \alpha_j^0 \alpha_i^0\rangle \varepsilon_j^0 \varepsilon_i^0\right)$$

$$\sum_{i=1}^{N} w_{ki}\left(\langle \alpha_j^P \alpha_i^P\rangle + \langle \alpha_j^0 \alpha_i^0\rangle \varepsilon_j^0 \varepsilon_i^0\right) = \langle \alpha_k^P \alpha_j^P\rangle \qquad (5.3.7)$$

There is now a system of equations of the form (5.3.7), one for each value of j. The unknown weights appear in all equations and can be found by e.g. Gaussian elimination. It should be noted that the left hand side of (5.3.7) involve correlations between prediction

errors at observation points and observation errors at the observation points whereas the right hand side has the prediction error correlations between observation points and analysis points (gridpoints). The knowledge of these statistical quantities is thus required in order to set up the system. In matrix form the coefficient matrix on the left hand side for the weights is called $\underline{\underline{M}}$ and the prediction errror vector on the right hand side $\underline{P}$. $\underline{\underline{M}}$ only involves observation points and can be split into its prediction error part $\underline{\underline{P}}$ and observation error part $\underline{\underline{O}}$. On the right hand side, the vector $\underline{P}$ involves both observation and grid points. In matrix form (5.3.7) can be written:

$$\underline{w}_k^T \underline{\underline{M}} = \underline{P}_k \quad ; \quad \underline{\underline{M}} = \underline{\underline{P}} + \underline{\underline{O}}$$

The weights can also be found by inverting M:

$$w_k = \underline{\underline{M}}^{-1} \underline{P}_k \tag{5.3.8}$$

Using these values in (5.3.6) for the weights gives also a value of the estimated interpolation error:

$$\varepsilon_k^{i\,2} = 1 - 2\, \underline{w}_k^T \underline{P}_k + \underline{w}_k^T \underline{\underline{M}}\, \underline{w}_k \tag{5.3.9}$$

Again it must be emphazised that these errors are the difference between the analysed state and the best possible state that we could obtain with the first guess model. As the number of observations increase to infinity this error goes to zero. It is not to say that we have determined the atmosphere exactly, only that we have fitted the atmosphere on the resolved scales exactly in a least square sense if all such cases are averaged and the statistics are the right statistics for this sample.

5.4 Simple example of O.I.

To see a very basic property of O.I. the effect of analysing only one observation is considered. The $\underline{\underline{M}}$ matrix has then a particularily simple form:

$$\underline{\underline{M}} = \underline{\underline{P}} + \underline{\underline{O}} = 1 + \varepsilon^{o\,2}$$

whereas the vector P is

$$\underline{P}_k = \mu_k$$

where μ_k is a correlation function. The inverse of $\underline{\underline{M}}$ and weight of the observation are

$$\underline{\underline{M}}^{-1} = 1/(1+\varepsilon^{o\,2}) \quad ; \quad w_k = \mu_k/(1+\varepsilon^{o\,2})$$

The analysis is given by

$$\frac{b \cdot \mu_k}{1+\varepsilon^{o\,2}}$$

where b is the normalised departure of the observation. At the observation point the analysed value is

$$\frac{b}{1+\varepsilon^2}$$

It can be seen that the analysed value is always smaller than the observed one. How much depends on the ratio between perceived observation error and prediction error. The interpolation to other points will again result in smaller values depending on the correlation function (<1).

In general O.I. can be viewed as a two-step procedure. First analysis at observation points meaning filtering of the data. Then interpolation to the gridpoints of the filtered data. These are indeed the two purposes of O.I. as described by Gandin (1963).

6. MULTI-VARIATE ANALYSIS

6.1 Introduction

The prediction error correlations <a a> in P can be considered as correlations between only the same variable at different points i and j. In this case the analysis is called uni-variate since only one variable is analysed using only one observed variable. When more than one observation variable is available and/or more than one analysis variable is required the analysis has to be performed several times with one variable at a time. This is the normal way of treating fields which are independent of each other or are difficult to relate. Examples of these are surface parameters like snowdepth, low-level wind and surface temperature. The correlations in P can also refer to cross-correlations between different variables at i and j if more than one observed variable is used when setting up the matrices P and C. Also the vector P contain such cross-correlations (between observations and grid-points) due to different observed variables but also due to the possibility of analysing several variables simultaneously. This is called multi-variate analysis and is often done for upper-air parameters like wind components, geopotential heights at pressure levels and pressures at geometric heihts. These are related to a good degree of aproximation through geostrophy and non-divergence.

In order for the O.I. analysis to produce reasonable results, the cross- correlations must satisfy the same rule as for the uni-variate correlations, namely that the P matrix is positive definite. This is a fundamental property of any correlation matrix and therefore care has to be taken when setting up such a matrix. Usually one models correlations as well-known function of distance and vertical separation. These functions should be such that they always produce positive definite correlation matrices and this can be ensured if they have a postive spectrum. $Exp(-0.5x^2/b^2)$ is such a function which has been widely used.

There are of course also more interesting constraints on the cross-correlations. The idea behind the multi-variate system is that we can make use of our knowledge of the physical and statistical relationships between the different variables; otherwise it would make no sense to try to analyse different variables simultaneously.

The geostrophic relationship is the most important example of such relationships to be utilized. The v-component and height cross-covariance can be derived as follows.

$$v = \frac{g}{f}\frac{\partial}{\partial x_j} z_j$$

where g is the constant of gravity and f the Coriolis parameter

$$\langle z_i v_j \rangle = \langle z_i \frac{g}{f}\frac{\partial z_j}{\partial x_j} \rangle = \frac{g}{f}\frac{\partial}{\partial x_j}\langle z_i z_j \rangle$$

The height-wind covariances can thus be derived from a model of only

height-height covariances. Note that the reasoning here has been confined to covariances and not correlations as used in the previously derived O.I. equations. When converting to correlations the geostrophic relationship should be used again in order to get the consistent normalisation factors E for heights and winds.

$$< V_i V_j > = \left(\frac{g}{f}\right)^2 \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} < z_i z_j > \; ; \; < z_i z_j > = e^{-0.5\,\frac{(x_i - x_j)^2}{b^2}} \cdot E_z^2$$

$$\frac{\partial}{\partial x_j} < z_i z_j > = \frac{x_i - x_j}{b^2} e^A E_z^2 \; ; \; \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} < z_i z_j > = \left[\frac{-(x_i - x_j)^2}{b^4} + \frac{1}{b^2}\right] e^A E_z^2$$

$$\lim_{x_i \to x_j} < V_i V_j > = \frac{1}{b^2}\left(\frac{g}{f}\right)^2 E_z^2 = E_v^2 \Rightarrow E_v = \frac{E_z}{b} \cdot \frac{g}{f}$$

Provided the observations are normalized in this way and correlations are modelled geostrophically the resulting analysis increments will satisfy the geostrophic relationship.

6.2 The full covariance model

All the auto-correlations and crosscorrelations together with the normalistations required are best derived from a model of prediction error covariances of geopotential, streamfunction and velocity potential (and their cross-covariances) following Daley (1985).

$$< \Phi \Phi > = E_\Phi^2 V_\Phi \Pi_{\Phi\Phi}(r) \qquad (6.2.1\,a)$$

$$< \Psi \Psi > = E_\Psi^2 V_\Psi \Pi_{\Psi\Psi}(r) \qquad (\quad b)$$

$$< \chi \chi > = E_\chi^2 V_\chi \Pi_{\chi\chi}(r) \qquad (\quad c)$$

$$< \Phi \Psi > = \mu E_\Phi E_\Psi V_{\Phi\Psi} \Pi_{\Phi\Psi}(r) \qquad (\quad d)$$

$$< \chi \Psi > = \gamma E_\chi E_\Psi V_{\chi\Psi} \Pi_{\chi\Psi}(r) \qquad (\quad e)$$

$$< \chi \Phi > = \eta E_\chi E_\Phi V_{\chi\Phi} \Pi_{\chi\Phi}(r) \qquad (\quad f)$$

Here homogenity and isotropy have been assumed since there is no angular and spatial dependence in the functions Π. The geopotential and streamfunction covariances usually have the same horizontal structure $\Pi(r)$ in order to facilitate the geostrophic constraint for the analysis via the geostrophic coupling constant μ in (6.2.1d). Another simplification is that the cross-covariances between velocity potential and streamfunction as well as between velocity potential and geopotential usually are set to zero since there is evidence that these terms are very small (see e.g. Hollingsworth and Lonnberg (1986)). The prediction error covariance model has been assumed to be separable in one horizontal part (Π) and another vertical part described by the various funtions V above.

The horizontal wind covariances can then be derived by differentiating the streamfunction and velocity potential covariances using Helmholtz's theorem :

$$\mathbf{v} = \nabla\chi + \mathbf{k}\times\nabla\psi \qquad (6.2.2)$$

or for the two horizontal wind components

$$u = \frac{\partial\chi}{\partial x} - \frac{\partial\psi}{\partial y} \quad ; \quad v = \frac{\partial\chi}{\partial y} + \frac{\partial\psi}{\partial x} \qquad (6.2.3-4)$$

We can then derive correlations between wind components at positions i and j assuming no horizontal variation in the error variances.

$$\langle u_i v_j\rangle = \left\langle \left(-\frac{\partial}{\partial y_i}\psi_i + \frac{\partial}{\partial x_i}\chi_i\right)\left(\frac{\partial}{\partial x_j}\psi_j + \frac{\partial}{\partial y_j}\chi_j\right)\right\rangle =$$

$$= -\frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j}\langle\psi_i\psi_j\rangle + \frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j}\langle\chi_i\chi_j\rangle -$$

$$- \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}\langle\psi_i\chi_j\rangle + \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\langle\chi_i\psi_j\rangle =$$

$$= -E_\psi^2\frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j}T_{\psi\psi} + E_\chi^2\frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j}T_{\chi\chi} +$$

$$+ E_\psi E_\chi\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} - \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}\right)T_{\psi\chi} \qquad (6.2.5)$$

$$\langle u_i u_j\rangle = \left\langle \left(-\frac{\partial}{\partial y_i}\psi_i + \frac{\partial}{\partial x_i}\chi_i\right)\left(-\frac{\partial}{\partial y_j}\psi_j + \frac{\partial}{\partial x_j}\chi_j\right)\right\rangle =$$

$$= \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}\langle\psi_i\psi_j\rangle + \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\langle\chi_i\chi_j\rangle -$$

$$- \frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j}\langle\psi_i\chi_j\rangle - \frac{\partial}{\partial y_j}\frac{\partial}{\partial x_i}\langle\psi_j\chi_i\rangle =$$

$$= E_\psi^2\frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}T_{\psi\psi} + E_\chi^2\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}T_{\chi\chi} -$$

$$- E_\psi E_\chi\left(\frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j} + \frac{\partial}{\partial y_j}\frac{\partial}{\partial x_i}\right)T_{\psi\chi} \qquad (6.2.6)$$

$$\langle v_i v_j\rangle = \left\langle \left(\frac{\partial}{\partial x_i}\psi_i + \frac{\partial}{\partial y_i}\chi_i\right)\left(\frac{\partial}{\partial x_j}\psi_j + \frac{\partial}{\partial y_j}\chi_j\right)\right\rangle =$$

$$= \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\langle\psi_i\psi_j\rangle + \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}\langle\chi_i\chi_j\rangle +$$

$$+ \frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j}\langle\psi_i\chi_j\rangle + \frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j}\langle\psi_j\chi_i\rangle =$$

$$= E_\psi^2\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}T_{\psi\psi} + E_\chi^2\frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}T_{\chi\chi} + E_\psi E_\chi\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j}\right.$$

$$\left. + \frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j}\right)T_{\psi\chi} \qquad (6.2.7)$$

and if we assume that the crosscorrelations $M_{\psi x}=0$

$$\langle u_i u_j \rangle = E_\psi^2 \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j} M_{\psi\psi} + E_x^2 \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} M_{xx} \qquad (6.2.8)$$

$$\langle v_i v_j \rangle = E_\psi^2 \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} M_{\psi\psi} + E_x^2 \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j} M_{xx} \qquad (6.2.9)$$

$$\langle u_i v_j \rangle = -E_\psi^2 \frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j} M_{\psi\psi} + E_x^2 \frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j} M_{xx} \qquad (6.2.10)$$

and if we assume isotropy; i.e.

$$M = M(r_{ij}) \quad ; \quad r_{ij} = \sqrt{(x_i-x_j)^2 + (y_i-y_j)^2}$$

then the partial derivatives can be transformed to radial derivatives only

$$\begin{cases} \dfrac{\partial}{\partial x_i} = \dfrac{\partial r}{\partial x_i}\dfrac{\partial}{\partial r} = \dfrac{x_i-x_j}{r}\dfrac{\partial}{\partial r} \quad ; \quad \dfrac{\partial}{\partial y_i} = \dfrac{y_i-y_j}{r}\dfrac{\partial}{\partial r} \\ \dfrac{\partial}{\partial x_j} = -\dfrac{(x_i-x_j)}{r}\dfrac{\partial}{\partial r} \quad ; \quad \dfrac{\partial}{\partial y_j} = -\dfrac{(y_i-y_j)}{r}\dfrac{\partial}{\partial r} \end{cases}$$

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} = \frac{\partial}{\partial x_i}\left(-\frac{(x_i-x_j)}{r}\right)\frac{\partial}{\partial r} - \frac{(x_i-x_j)}{r}\frac{\partial}{\partial x_i}\frac{\partial}{\partial r} =$$

$$= \left[-\frac{1}{r} + \frac{(x_i-x_j)^2}{r^3}\right]\frac{\partial}{\partial r} - \frac{(x_i-x_j)^2}{r^2}\frac{\partial^2}{\partial r^2}$$

$$\frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j} = \frac{\partial}{\partial y_i}\left(-\frac{(y_i-y_j)}{r}\right)\frac{\partial}{\partial r} - \frac{(y_i-y_j)}{r}\frac{\partial}{\partial y_i}\frac{\partial}{\partial r} =$$

$$= \left[-\frac{1}{r} + \frac{(y_i-y_j)^2}{r^3}\right]\frac{\partial}{\partial r} - \frac{(y_i-y_j)^2}{r^2}\frac{\partial^2}{\partial r^2}$$

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j} = \frac{\partial}{\partial x_i}\left(-\frac{(y_i-y_j)}{r}\right)\frac{\partial}{\partial r} - \frac{(y_i-y_j)}{r}\frac{\partial}{\partial x_i}\frac{\partial}{\partial r} =$$

$$= \frac{(y_i-y_j)(x_i-x_j)}{r^3}\frac{\partial}{\partial r} - \frac{(x_i-x_j)(y_i-y_j)}{r^2}\frac{\partial^2}{\partial r^2}$$

$$\frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j} = \frac{\partial}{\partial y_i}\left(-\frac{(x_i-x_j)}{r}\right)\frac{\partial}{\partial r} - \frac{(x_i-x_j)}{r}\frac{\partial}{\partial y_i}\frac{\partial}{\partial r} =$$

$$= \frac{(x_i-x_j)(y_i-y_j)}{r^3}\frac{\partial}{\partial r} - \frac{(x_i-x_j)(y_i-y_j)}{r^2}\frac{\partial^2}{\partial r^2}$$

inserting these expressions in Eq. 6.2.8-10 gives:

$$\langle u_i u_j \rangle = E_\psi^2 \left[-\left(1 - \frac{(y_i - y_j)^2}{r^2}\right) \frac{1}{r}\frac{\partial}{\partial r} - \frac{(y_i - y_j)^2}{r^2}\frac{\partial^2}{\partial r^2} \right] \eta_{\psi\psi} + E_\chi^2 \left[-\left(1 - \frac{(x_i - x_j)^2}{r^2}\right) \frac{1}{r}\frac{\partial}{\partial r} - \frac{(x_i - x_j)^2}{r^2}\frac{\partial^2}{\partial r^2} \right] \eta_{\chi\chi}$$

$$\langle v_i v_j \rangle = E_\psi^2 \left[-\left(1 - \frac{(x_i - x_j)^2}{r^2}\right) \frac{1}{r}\frac{\partial}{\partial r} - \frac{(x_i - x_j)^2}{r^2}\frac{\partial^2}{\partial r^2} \right] \eta_{\psi\psi} + E_\chi^2 \left[-\left(1 - \frac{(y_i - y_j)^2}{r^2}\right) \frac{1}{r}\frac{\partial}{\partial r} - \frac{(y_i - y_j)^2}{r^2}\frac{\partial^2}{\partial r^2} \right] \eta_{\chi\chi}$$

$$\langle u_i v_j \rangle = \frac{(x_i - x_j)(y_i - y_j)}{r^2} \left[\frac{1}{r}\frac{\partial}{\partial r} - \frac{\partial^2}{\partial r^2} \right] \left[-E_\psi^2 \eta_{\psi\psi} + E_\chi^2 \eta_{\chi\chi} \right]$$

From Eq. 6.2.11-13 it is worth noting a few things:

* $\langle vv \rangle$ can be found by rotating $\langle uu \rangle$ 90 degrees.
* Along the x-axis $\langle uu \rangle$ is:

$$-E_\psi^2 \frac{1}{r}\frac{\partial \eta_{\psi\psi}}{\partial r} - E_\chi^2 \frac{\partial^2 \eta_{\chi\chi}}{\partial r^2} \qquad (6.2.14)$$

* Along the y-axis $\langle uu \rangle$ is:

$$-E_\psi^2 \frac{\partial^2 \eta_{\psi\psi}}{\partial r^2} - E_\chi^2 \frac{1}{r}\frac{\partial \eta_{\chi\chi}}{\partial r} \qquad (6.2.15)$$

* $\langle uv \rangle$ is zero along both x- and y-axes.

6.3 The effect of including divergent correlations

The total wind prediction error variance Ev2 is divided into a divergent part (ν) and non-divergent part ($1-\nu$).

$$E_v^2 = E_\psi^2 + E_\chi^2$$

$$E_\chi^2 = \nu E_v^2$$

$$E_\psi^2 = (1-\nu) E_v^2$$

The effect on the wind correlations of increasing the divergent part of the wind forecast error can be seen in Figs. 1 and 2 (the same horizontal scale is employed for both stream function and velocity potential). They display the vector and scalar correlations with a single u-component situated at 40 N, 40 W. The correlations are exactly as they would be in the ECMWF analysis system with full spherical geometry is employed except that the constant large scale term (see HL) is omitted for clarity. The v-v correlations are not shown here; they can be found by just rotating the u-u correlations 90 degrees anti-clockwise (apart from some change in shape due to map factors) and the v-u correlation is always the same as u-v as can be deduced from Eqs. (6.2.5) and (6.2.6). As ν is introduced with a value of 0.1 the major effect is to reduce the cross-correlation u-v and the negative lobes of the u-u correlation. The combined effect for the vectors is a significantly reduced return flow as the top panels in Fig. 1 show. When ν=0.5 the auto-correlation becomes circular (Fig. 2) and the cross-correlation zero since the longitudinal and transverse correlations now are equal ((6.2.5) and(6.2.6)). (This is of course only true when the same horizontal structures F and G have been employed). With this choice of the analysis would be uni-variate in u and v and the definition of (6.2.3a-c) implies that equal amounts of vorticity and divergence are assumed for the first guess error. Lorenc (1979) pointed out this consequence of analysing u and v components separately. Finally, when using purely divergent structure functions with ν=1 (Fig. 2 right) we see that the u-u and u-v correlations are of the same shape as for the non-divergent case but they are turned anti-clockwise 90 degrees.

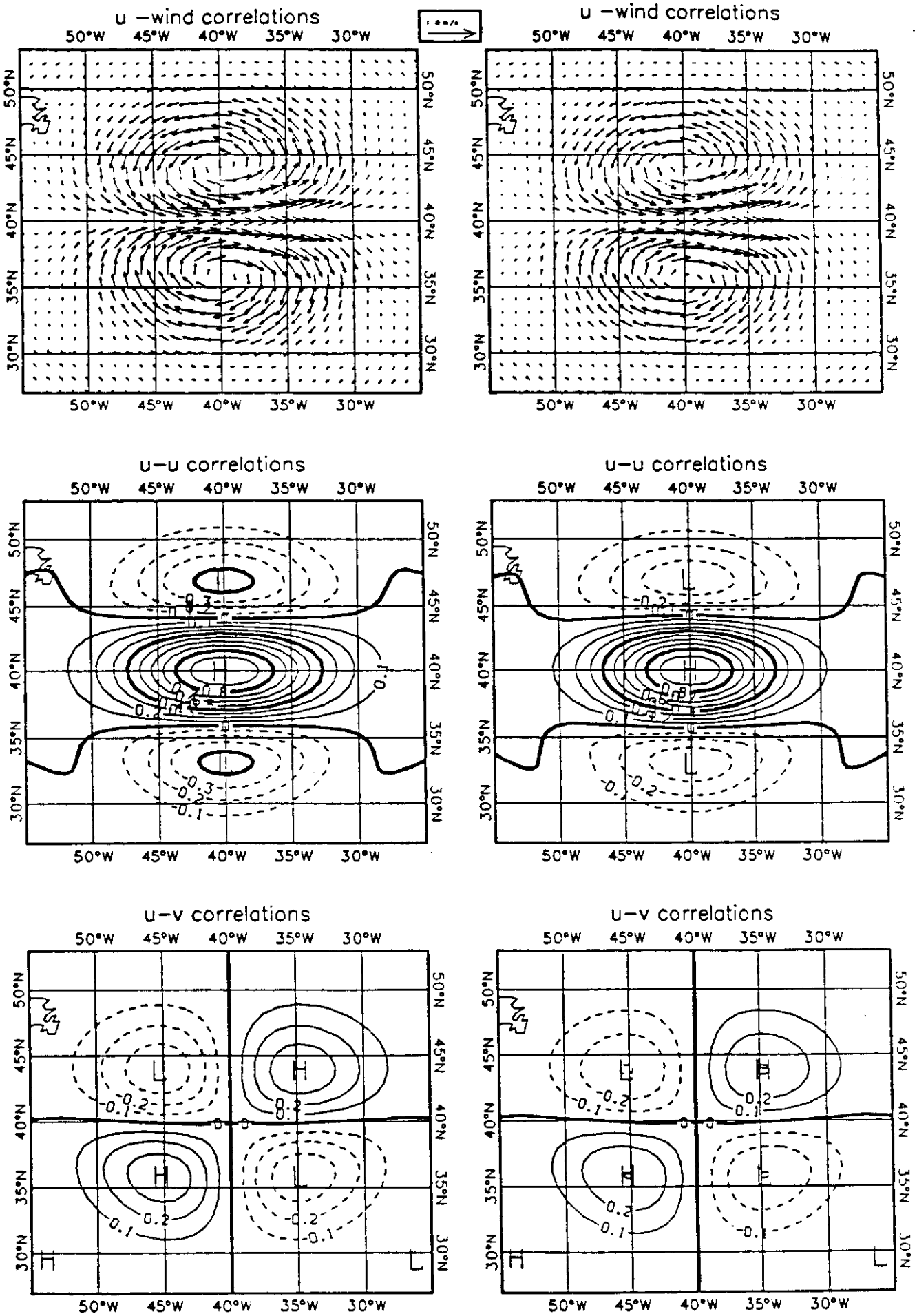


Figure 1. Horizontal correlations between a u wind component at 40N, 40W and wind vectors, u and v components for ny=0 (left) and ny=0.1 (right).

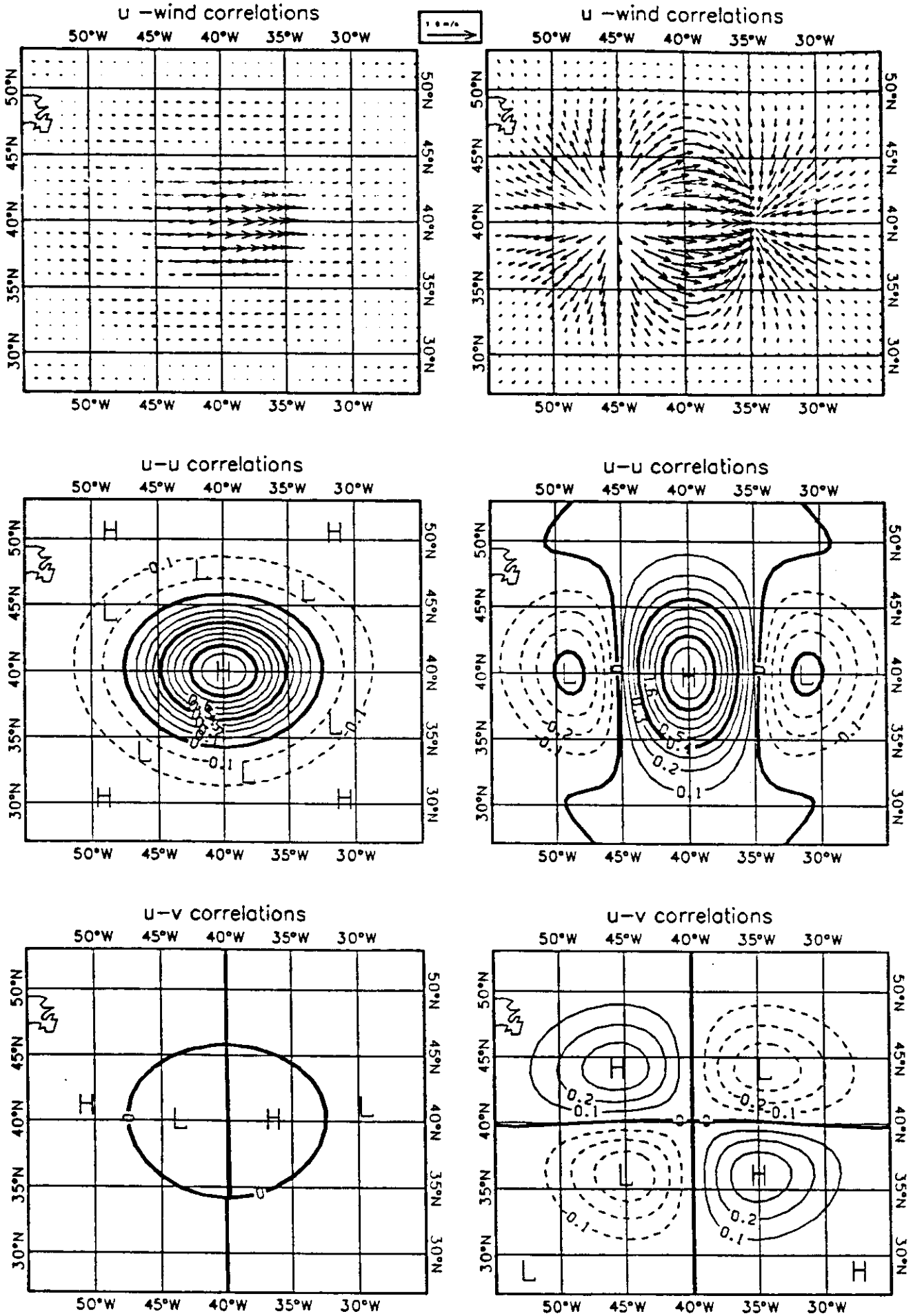


Figure 2. As Fig. 1 but for ny=0.5 (left) and ny=1.0 (right).

7. THE STATISTICAL STRUCTURE OF THE SHORT RANGE FORECAST ERRORS

7.1 Introduction

A vital component of the O.I. system is the choice of the statistics used to model all the correlations necessary in the M and P matrices. This section describes how this has been done for prediction errors (P) for the ECMWF system. The work has been done by Hollingsworth and Lonnberg (1986) and the following description is a summary of their results.

7.2 Isotropic covariances

In the previous section it was shown how the covariances for the two horizontal wind components can be derived. Also the expressions for ‹uu› and ‹vv› along the x and y axes were given there.

$$\langle UU\rangle_x + \langle UU\rangle_y = -\left(\frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2}\right)\left(E_\psi^2 M_{\psi\psi} + E_\chi^2 M_{\chi\chi}\right) \quad (7.2.1)$$

$$\langle UU\rangle_x - \langle UU\rangle_y = -\left(\frac{1}{r}\frac{\partial}{\partial r} - \frac{\partial^2}{\partial r^2}\right)\left(E_\psi^2 M_{\psi\psi} - E_\chi^2 M_{\chi\chi}\right) \quad (7.2.2)$$

Now, if a large set of statistics of departures between the 6-hour first guess forecast values and observed values is formed covariances can be computed for all pairs of stations. For each such pair the coordinate system can be rotated so that the x-axis is along the line joining the two stations and the y-axis orthogonal to this line. The local wind components of the departures are then transformed to the new system and the system (7.2.1-2) above can be used to solve the correlation functions for the rotational and divergent part of the wind given covariances of along the x and y directions for ‹uu›. Since differential operators are involved here a model function of distance is required and the coefficients or scale lengths in this function can be determined for the two parts of the wind field. The model used here for the stream function and velocity potential is a series of Bessel functions

$$M(r) = \sum_n A_n J_0(k_n r/D)$$

In practice, to get a larger sample, the assumption of homogeneiaty is is also used and covariances between different pairs of stations with similar separation distances are grouped together. Bins of 100 km have been employed here.

The coefficients of the Bessel functions can be fitted allowing differnent truncations. Fig. 3 shows the resulting fits to the sum of ‹uu› and ‹uu› in (7.2.1-2) for various number of terms in the series.

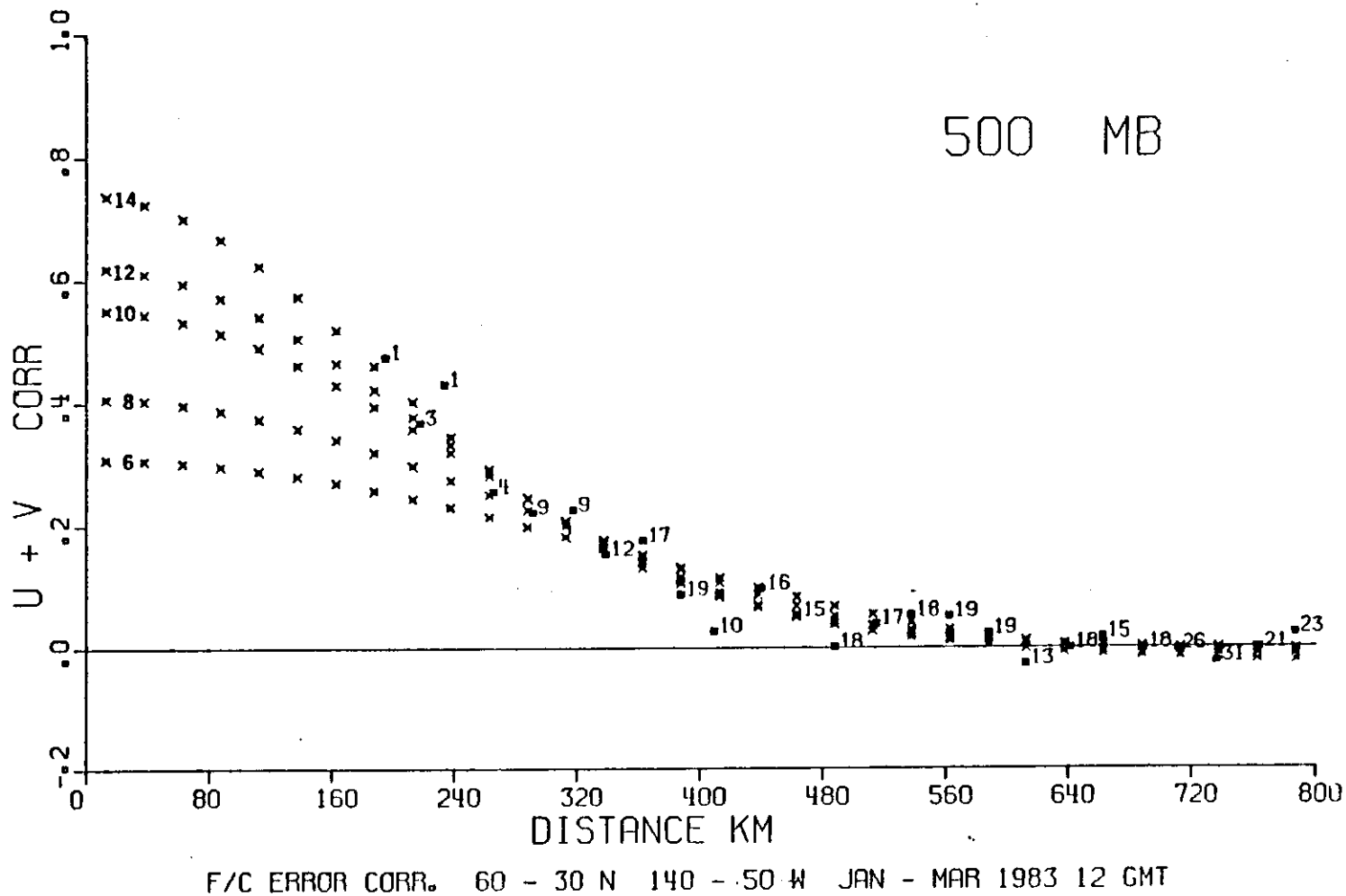


Fig. 3 The variation,at 500mb, of the <ℓ,ℓ>+<t,t> correlation with station separation: The squares show the empirically determined average value for each 25km 'bin', together with the number of station pairs in that bin. Only data for separations of 800km or less is shown. All the data out to 3000km was used in the least squares procedure to determine the five fitting curves (x) with truncations of 6, 8, 10, 12 and 14 synoptic scale terms as indicated.

It is seen that allowing for only few terms gives a very smooth function and allowing for more gives more structure and implies a higher resolution. An important point is the separation of observation errors and prediction errors. At zero separation the prediction error correlations in (7.2.1-2) , after dividing by the variances, are equal to one (can be verified by e.g. using the exponential function of previous section). When extrapolating the curves in Fig. 3 to zero it is clear that the interception with the y-axis is far from one. This discrepancy is the perceived observation error. It is a random horizontally uncorrelated error which is present in the covariance calculations since the first guess departures contain both instrumental errors and representativeness errors by the forecast model. Furthermore, it can be seen that increasing the number of terms in the Bessel representation reduces this perceived observation errors since smaller scales are resolved. The choice of the number of terms should however be consistent with the best fit at all separations considered and this represents the actual resolution of the forecast model.

Vertical correlations can be computed by fitting correlations for the wind shear between two levels

$$\Delta U = U^n - U^m$$

The wind shear error can be divided into a prediction error and an observation error part

$$\Delta U = \Delta p + \Delta b$$

Then forming the covariance gives

$$\langle \Delta U_i^n, \Delta U_j^m \rangle = \langle \Delta p_i + \Delta b_i , \Delta p_j + \Delta b_j \rangle =$$

$$= \langle \Delta p_i \Delta p_j \rangle + \langle \Delta b_i \Delta b_j \rangle \delta_{ij}$$

assuming that prediction and observation errors are uncorrelated.

7.3 Results

It has already been mentioned that the truncation of the Bessel series must be compatible with the resolution of the forecast model. The statistics presented here is based on a T63 spectral forecast model and the maximum resolution for that model corresponds nearly to 10 terms in (7.2.2). Going beyond that gives instability in the calculation of curve fits. With this truncation the functions can be fitted and the zero distance interception used to partion the error into prediction and observation errors. This is repeated for all levels and the perceived errors are shown in Fig. 4. It is seen that two errors are similar in magnitude at lower levels and that the prediction error then is somewhat larger at upper levels. The total error has its maximum around the typical jet stream levels as can be expected.

Fig. 5 and 6 show the prediction error correlation for $\langle uu \rangle_x$ and $\langle uu \rangle_y$ in (6.2.14-15) (also called longitudinal and transverse components). The latter shows a much sharper correlation with a negative lobe at some distance. This is a result of the dominance of the non-divergence in the data as can be seen by comparing with Fig. 1 with no or little divergence in previous section. The negative lobe in Fig. 1 is due to the return flow imposed by the non-divergent constraint. At some levels this negative part may however be damped by a constant large scale prediction error which has to be separated out before drawing any conclusions about the amount of divergence present. The partioning into rotational and divergent error is shown in Fig. 7 and the divergent part is significantly smaller than the non-divergent at all levels. The ratio between divergent error and total error is around 0.30-0.35 as Fig. 8 shows. In variance terms it means that about 10 % of the total wind errror is explained by divergent structures and this number can be used for parameterising them in a covariance model.

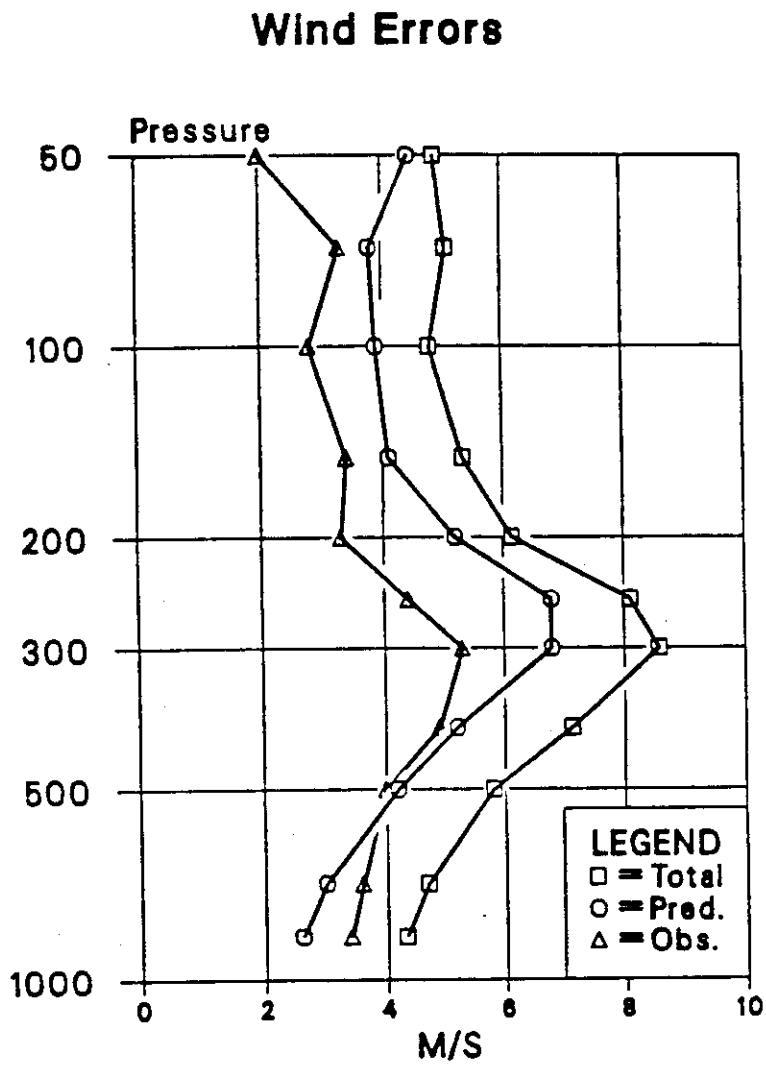


Fig. 4 The vertical variation of the perceived wind forecast errors (Total), together with the corresponding profiles of the prediction (Pred) and observation (Obs) error.

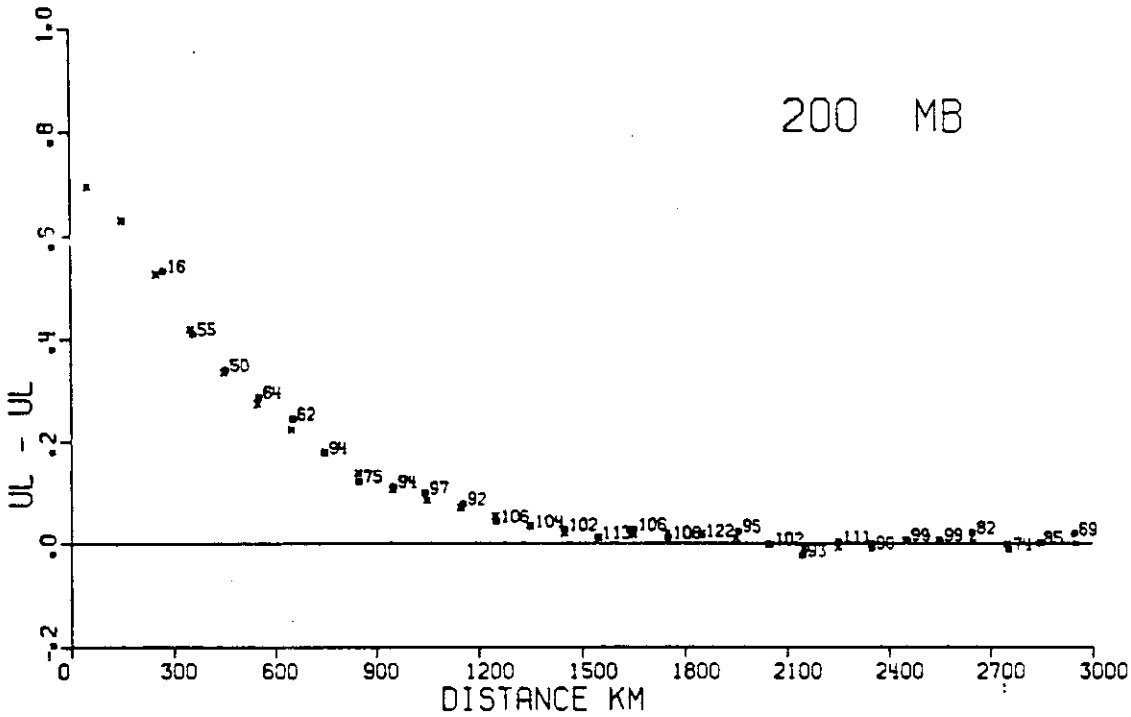


Fig. 5 The variation, at 200mb, of the $\langle \ell, \ell \rangle$ or longitudinal correlation with station separation: The squares show the empirically determined average value for each 100km 'bin', together with the number of station pairs in that bin. All the data out to 3000km was used in the least squares procedure to determine the fitting curve (x) with a truncation of 10 synoptic scale terms.

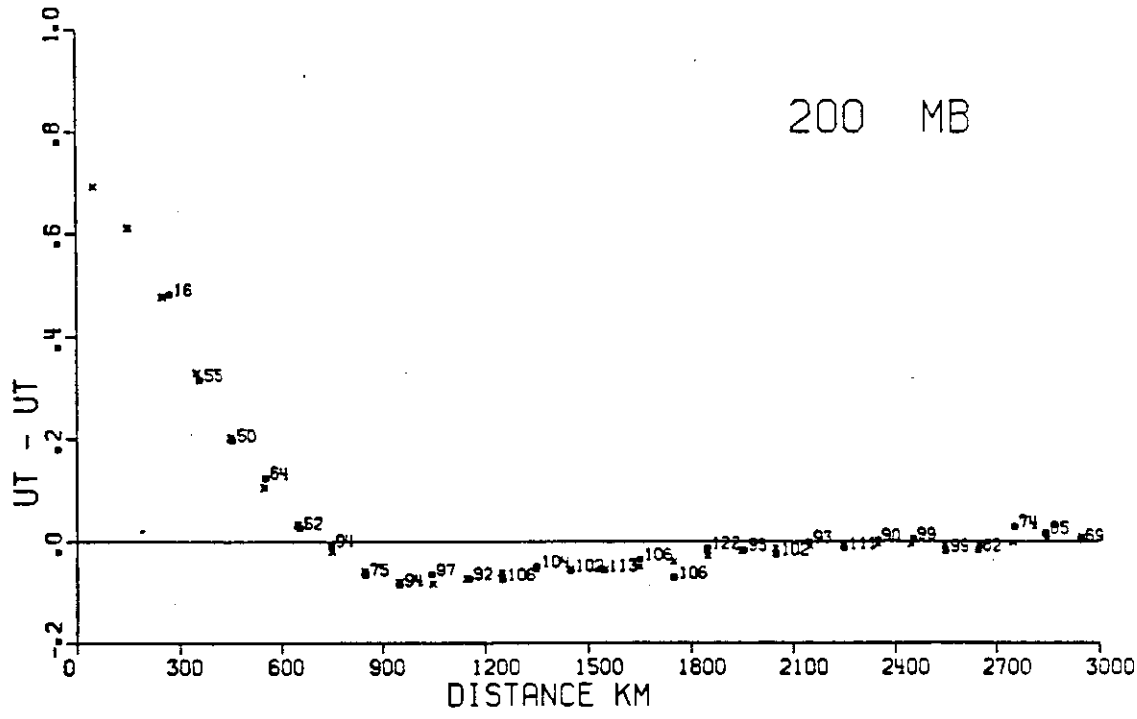


Fig. 6 The variation at 200mb of the $\langle t, t \rangle$ or transverse correlation with station separation: The squares show the empirically determined average value for each 100km 'bin', together with the number of station pairs in that bin. All the data out to 3000km was used in the least squares procedure to determine the fitting curve (x) with a truncation of 10 synoptic scale terms.

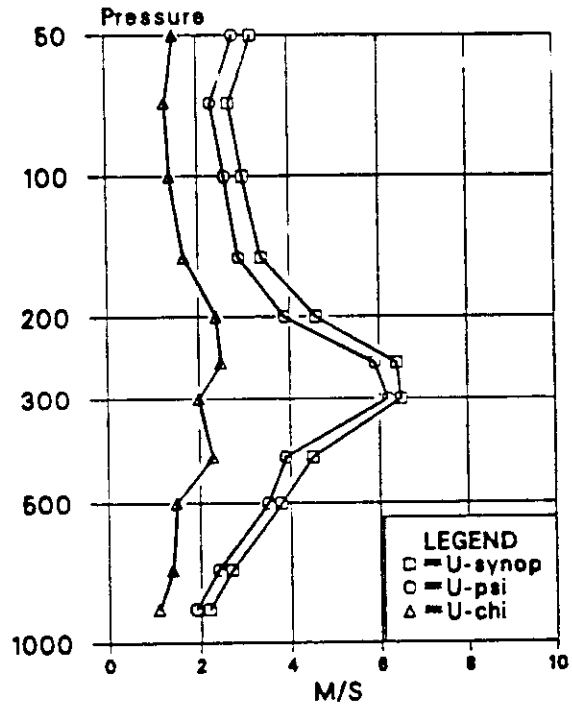


Fig. 7 The vertical profile of the rms synoptic-scale prediction error for vector wind (U-synop), together with the rms non-divergent (U-psi) and rms divergent (U-chi) contributions to the total. The sum of the variances of the two contributions gives the variance of the synoptic-scale prediction error.

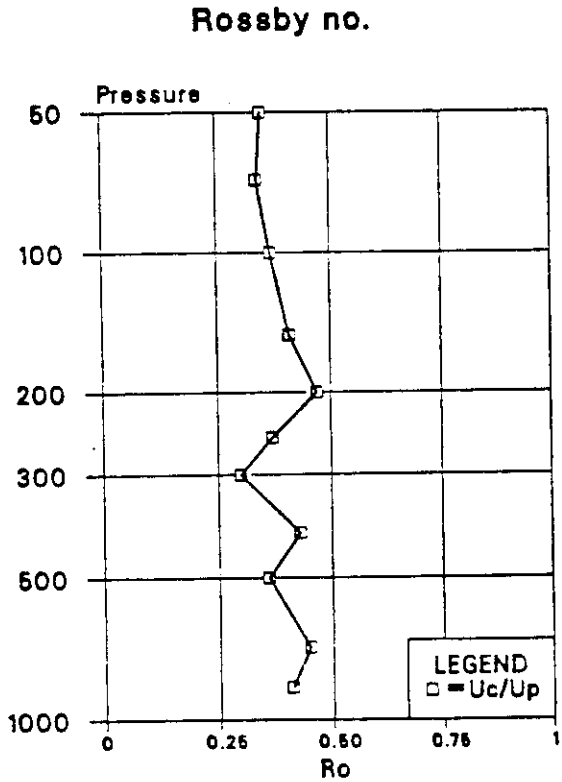


Fig. 8 The vertical profile of the Rossby number of the wind forecast errors, defined as the ratio of the rms divergent wind error to the rms prediction error, which includes both large-scale and synoptic-scale terms.

Finally the vertical correlations are derived and these are shown in Fig. 9 and 10 for the rotational and divergent part of the wind error respectively. For the rotational part, the correlations decay at increased vertical separation but remain mainly positive close to zero. The divergent error on the other hand has a negative part at some separation and then returns to zero. These curves have to be born in mind when parameterizing vertical correlations with some functional expression.

The technique described here can be repeated for the height observations with the simplification of only dealing with a scalar variable. This was done by Lonnberg and Hollingsworth (1986). Fig. 11 shows the fit of the Bessel series to height departures. It is of similar shape to Fig. 5 for the longitudinal part of the wind correlations which can expected of the divergence is low and geostrophy high. One differnce is though that the curve does not go to zero at large separation. This means that the constant term in the height errors is larger for heights than for winds. This is also shown in Fig. 12b at different levels. The partition into prediction and perceived observation error is displayed in Fig. 12a. It can be seen that the prediction error in a 6-hour forecast is really quite small (on average) and this is why it is such a powerful tool in data checking and monitoring as well as providing a very good first guess field for an analysis.

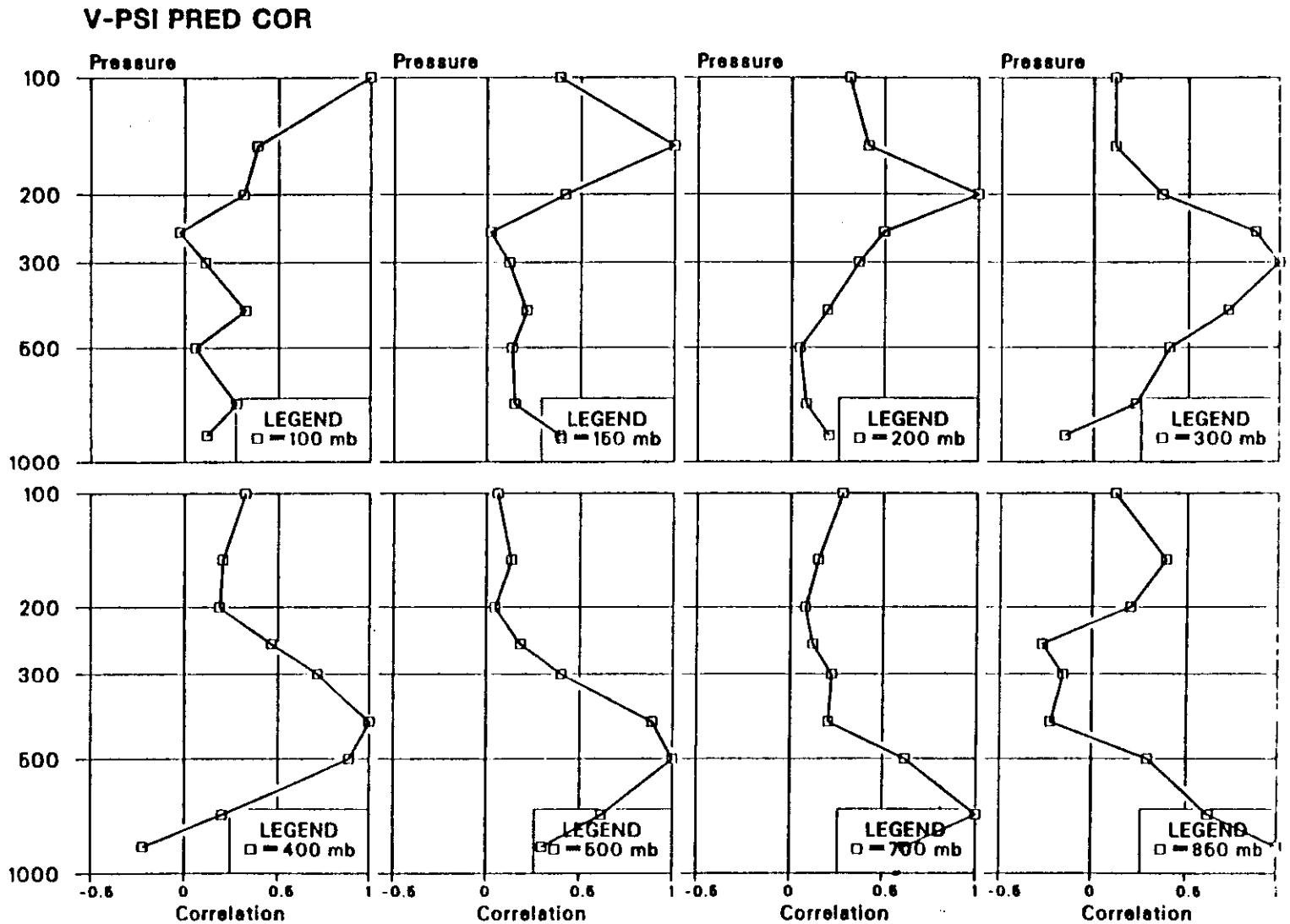


Fig. 9 Non-divergent wind prediction error vertical correlations for a selected set of standard levels, indicated in the legend of each frame of the plot. The plots corresponds to particular columns of the vertical correlation matrix for non-divergent wind.

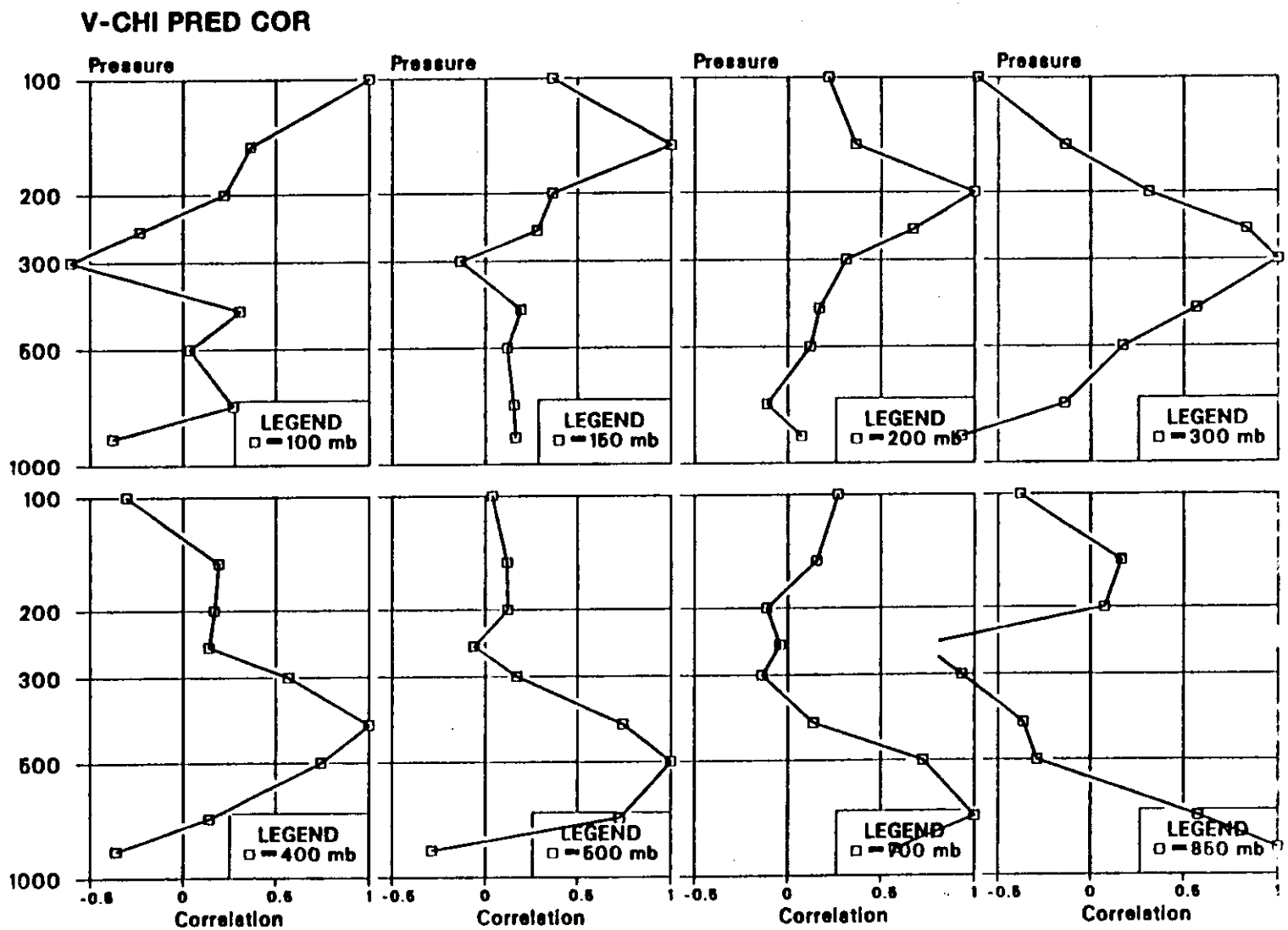


Fig. 10 As figure for the first term in the expansion of the divergent wind forecast errors.

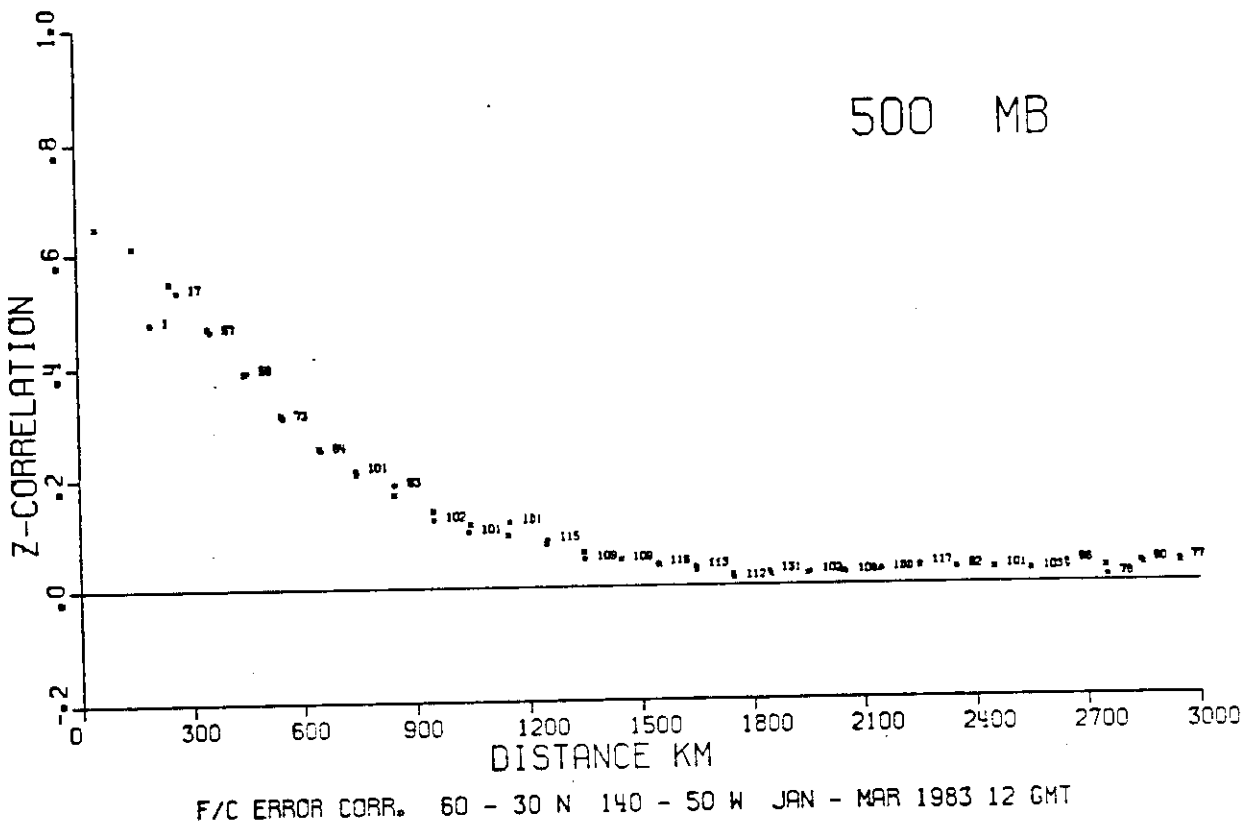


Fig. 11 The correlation of 500mb height forecast errors as a function of station separation. The empirical data, averaged over 'bins' of 100km, is shown by the squares. The figures indicate the number of station pairs in each bin. The smooth curve (x) is obtained by a least squares fit to the empirical data; the truncation is 8 terms in the synoptic-scale component

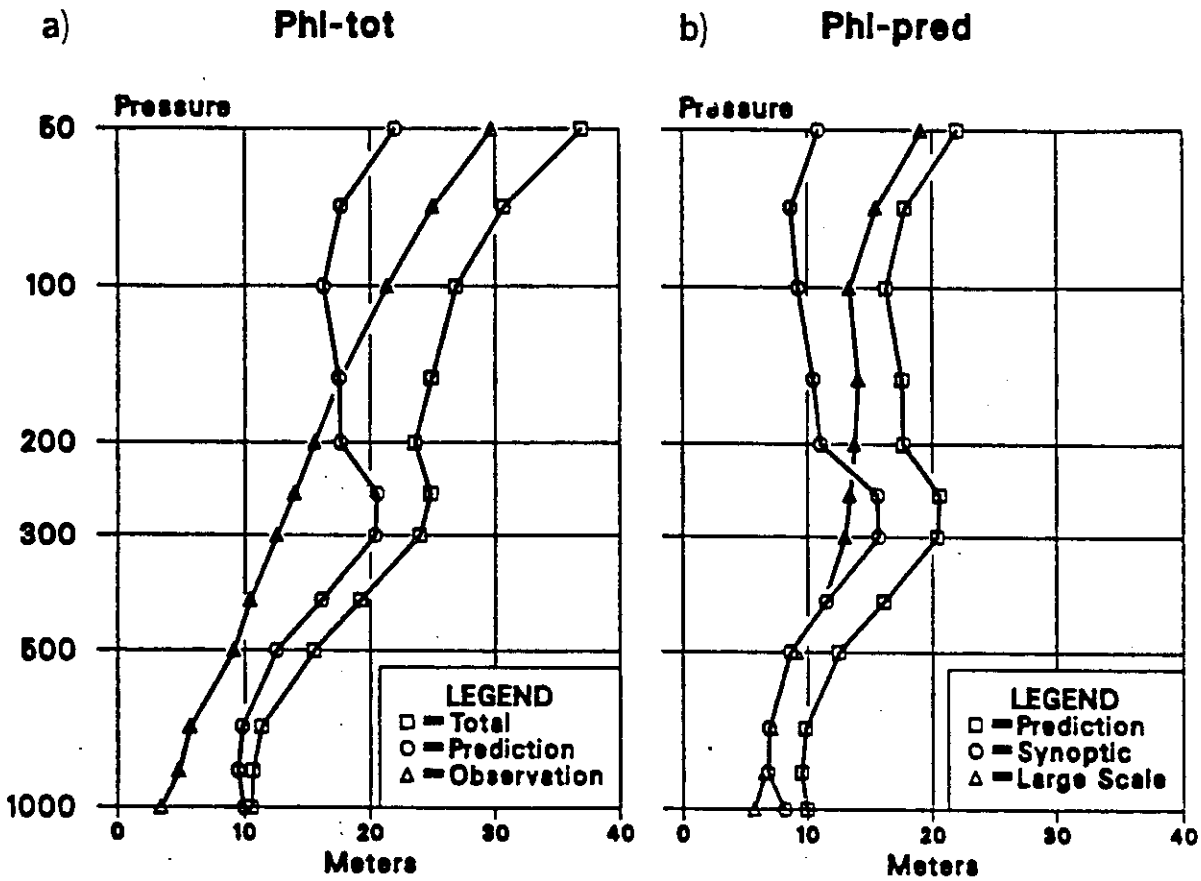


Fig. 12a: Vertical profiles of the total, or perceived, forecast error of height, together with the contributions to this error from the the prediction error, and the observation error. The unit is metre.

12b: Vertical profiles of the prediction error (copied from 2a) and of the contributions of the synoptic-scale and large-scale components to the prediction error. The sum of the squares of the components gives the square of the prediction error.

8. THE ECMWF ANALYSIS SYSTEM

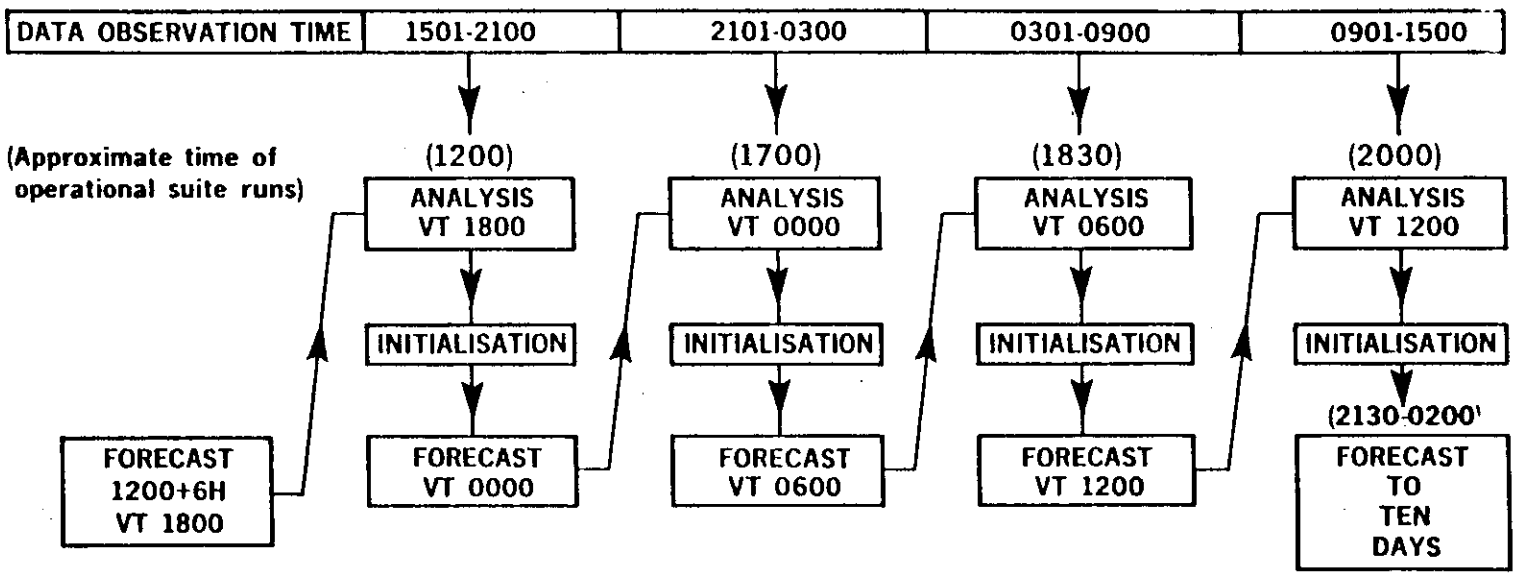
8.1 General properties

The ECMWF data assimilation system is an intermittent system. All observations within a 6-hour slot (-179,180 minutes) are grouped together in one big file. Where possible pressure corrections are done for SYNOP and SHIP data to correct for the time difference. Also some increase of the perceived observation errors is made for off-time data. Otherwise the data are used just as if they were all made at the analysis time. The first guess is from a 6-hour forecast made with the same forecast model as is used for the 10-day daily forecasts (see Fig. 13). The analysis is succeeded by a normal-mode initialisation and then another 6-hour forecast will produce the first guess for the next analysis and so on.

The mass and wind upper air analysis is a 3-dimensional optimum interpolation (O.I.) analysis multivariate in heights, wind components and satellite thicknesses. These are the input variables to the analysis and the output variables are heights and winds on the model coordinates.

8.2 Use of observations and statistics

Relevant input information is taken from all observation types. From SYNOP's and SHIP's the pressures (or in some cases heights) are used everywhere except in the case of excessive extrapolation done at the station or from the first guess due to large difference between model and real orography. Winds are only used from SHIP and low level tropical SYNOP's. Over land there is a severe difficulty to use SYNOP winds due to the inability of the model to produce realistic friction effects on the first guess winds. Pressures and winds are used from DRIBU's. Aircraft winds (AIREP and ASDAR) are also used. Temperature information from the these single-level observation types mentioned so far is not used.



OPERATIONAL DATA ASSIMILATION - FORECAST CYCLE

Fig. 13 The 6 hourly intermittent data assimilation.

From upper air multi-level observations the nearest available data to a standard pressure levels are selected (1000-850-700-500-400-300-250-200-150-100-70-50-30-20 and 10 hPa). These are the standard levels normally present in TEMP reports. Few however reach the top level at 10 hPa (about 31 km). Winds and heights from TEMP's are used. If a standard level height is missing it may be reconstructed from temperatures. PILOT's report only winds and they are also used at or near the standard levels. Satellite winds, SATOB's, are used from the GOES, METEOSAT and GMS satellites but not over land in the extra-tropics. Satellite thicknesses, SATEM/TOVS, are used but combined into 7 layers (1000-700, 700-500, 500-300, 300-100, 100-50, 50-30 and 30-10 hPa). They are only used above 100 hPa over land. Additionally there are Australian bogus observations, PAOBS, containing estimated pressures over the southern oceans and they are used as well.

The O.I. scheme relies, as mentioned earlier on various statistical information. The perceived observation errors for the observation types mentioned in Section 6 are shown in Table 1. It must be stressed again that they apart from the instrumental and rounding errors also contain the representativeness errors due to the imcompatabilities between the resolution of the model and scales on which the instrument samples. The perceived observation errors determine the relative weight given to various observation types and the weight given to the first guess. The mean 6-hour first guess errors (standard deviations) are seen at the bottom of Table 2 for the winter in the extra-tropics. It also contains the vertical first guess height correlations between standard levels used as they would be used to correlate e.g. the levels in a complete TEMP. Correlations between winds are the same in the extra-tropics due to the geostrophic constraint but both vary with season as well as the first guess errors do. In the tropics the wind error correlations are significantly sharper than for the heights. Also observation errors have vertical correlations for TEMP winds and heights as well ase for satellite thicknesses. The latter also have a horizontal observation error correlation.

The horizontal first guess error correlations are modelled using sums of Bessel functions with a component length scale of 500 km in the nothern hemisphere, 1000 km in the tropics and 600 km in the southern hemisphere extra-tropics (see Fig. 14).

Inst. Type	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
SONDE/PILOT WIND	2.2	2.5	2.6	3.1	3.7	3.8	3.3	3.0	2.8	2.4	2.4	2.4	2.5	3.1	3.5
SONDE GEOP.	5.0	5.4	6.0	9.4	11.6	13.8	14.2	15.2	18.2	21.4	25.2	29.8	31.2	38.1	50.0
SATEMS															
Clear path	-	11.8	11.0	14.6	11.8	14.2	-	22.5	-	23.8	16.8	17.3	24.5	-	52.4
Partly cloudy	-	13.1	12.2	16.2	13.1	15.8	-	25.0	-	26.4	18.7	19.2	27.2	-	58.2
Microwave	-	14.4	13.4	17.9	14.4	17.4	-	27.5	-	29.0	20.6	21.2	29.9	-	64.0
SATOBS															
GOES METEOSAT HIMAWARI	2.5	2.5	2.5	2.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
AIDS/ASDAR AIREP	3.5	3.5	3.5	3.5	3.5	4.0	4.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
PAOBS	32.0														
SYNOP/SHIP WIND	3.6														
SYNOP/SHIP GEOP.	7.0														
DRIBU GEOP.	14.0														

Table 1. The rms observation errors (units ms^{-1} and m). The SATEM thickness errors are given for layers with pressure at the top of the layer given in the first row.

WINTER Z CORRELATIONS(*1000) & ERRORS(M)

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
1000	1000	718	457	236	174	135	108	75	53	36	27	22	16	15	11
850	718	1000	813	481	353	260	200	126	82	55	41	34	25	24	18
700	457	813	1000	745	578	433	333	201	119	75	55	45	35	32	25
500	236	481	745	1000	907	748	607	377	207	115	79	62	47	44	35
400	174	353	578	907	1000	907	774	509	281	146	94	71	52	49	39
300	135	260	433	748	907	1000	928	669	384	193	115	81	57	52	41
250	108	200	333	607	774	928	1000	813	500	256	149	101	68	61	48
200	75	126	201	377	509	669	813	1000	767	436	252	159	96	81	62
150	53	82	119	207	281	384	500	767	1000	737	468	294	160	123	86
100	36	55	75	115	146	193	256	436	737	1000	805	561	308	218	132
70	27	41	55	79	94	115	149	252	468	805	1000	838	517	364	203
50	22	34	45	62	71	81	101	159	294	561	838	1000	758	562	311
30	16	25	35	47	52	57	68	96	160	308	517	758	1000	800	555
20	15	24	32	44	49	52	61	81	123	218	364	562	800	1000	750
10	11	18	25	35	39	41	48	62	86	132	203	311	555	750	1000
	9.6	9.6	9.6	12.4	15.4	19.6	20.4	19.1	17.0	16.3	17.6	22.0	28.0	37.0	50.0

Table 2. Vertical forecast error correlations (x1000), together with the error amplitudes at the levels (1000,850,.....,10 hPa). The errors are cosidered to be representative of continental areas. Both sets of are for winter extratropical heights. Units in m.

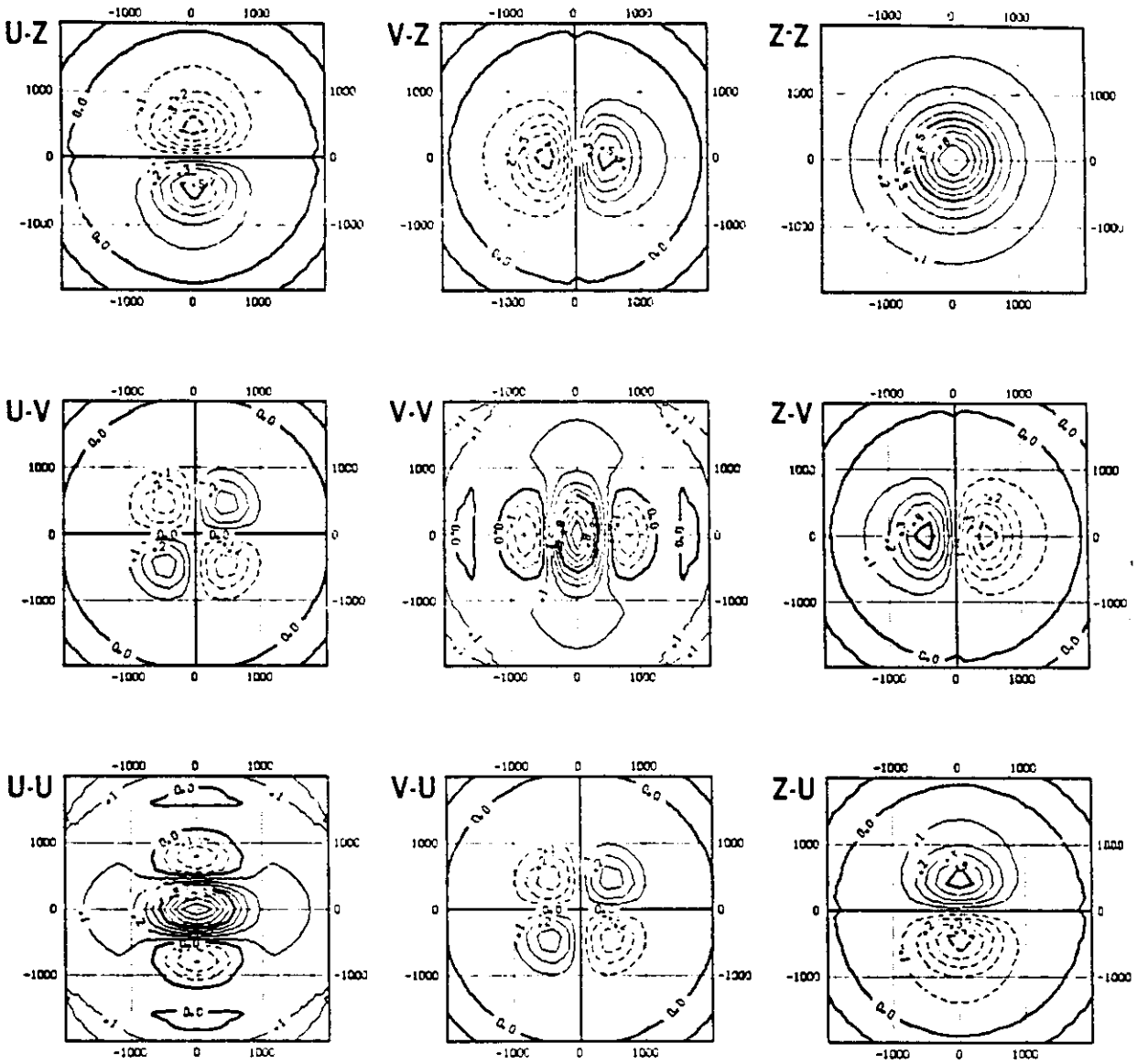


Fig. 14 Height and wind correlations, based on a bessel function model, as a function of separation (km). The component length scale is 500 km.

8.3 Computational organisation

The first guess is transformed from the model's spectral space and interpolated vertically to all data levels present in a model grid box. Then an horizontal bilinear interpolation from the four surrounding gridpionts is made. The thus constructed first guess value can now be compared with the observed value but first a normalisation takes place. The departure is divided by the estimated first guess error at the observation point. This is computed from the O.I. analysis errors from previous analysis cycle but inflated with time. All calculations are from here onwards done on normalized values. This makes it easier to compare departures from different regions and especially from different levels in the atmosphere.

A first guess quality control is made on the normalized departure. If it exceeds a value

$$\frac{(A^o - A^p)^2}{E^{p2}} > n^2 \left(\frac{E^{o2} + E^{p2}}{E^{p2}}\right)$$

the datum is flagged as incorrect or probably incorrect. A and A are the observed and first guess values and E and E the observation and first guess errors respectively. n has values of 4,6 or 8 empirically depending on observation types. If the departure is slightly lower the datum is considered to be possibly incorrect and if a number of subsequent levels are flagged these levels and higher ones will be rejected (not used) for TEMP's or the whole report may be rejected for a SATEM/TOVS. An additional check is done for the wind directions. If the wind speed exceeds 15 m/s the wind is rejected the direction differs from the first guess by more than 110 or 140 degrees (depending on level type).

Climatological and internal consistency checks have already been done at a pre-processing stage. Time consistency checks are only performed to a certain extent for SHIP's. If a SHIP with unique call sign has reported at least 3 times during the last 48 hours and has been flagged by the O.I. check (see below) on a majority of occasions it is blacklisted. A dynamical blacklist file is kept and updated operationally. Other observation types can at present (1988) only be manually blacklisted (exluded from analysis permanently). Such decisions are based on operational day to day monitoring as well as on monthly statistics.

The data processed so far are basically all the data from the observations and not all are used. The desired variables and levels from the reports are selected according to criteria mentioned earlier in Sect. 8.2. Redundant data are then removed and only the most appropriate are kept. TEMP winds has precedence over PILOT winds if they are for the same level and time. If there are more reports from the same station (SYNOP, SHIP and DRIBU) only the one closest to the analysis time is kept. This subset of observational data is then presented to the O.I. mass and wind analysis.

A further thinning is made of closely located SYNOP and SHIP data.

If there are three or more inside a model box on the Gaussian grid of size 1.125 by 1.125 degrees latitude and longitude they are combined into one super-observation. This is done through a local O.I. analysis (preceeded by O.I. data checking) with a modification so as to take no information from the first guess. The position of the super-observation is chosen as the average of the constituting observations. Those are no longer used in the analysis.

The next step is the O.I. data checking. The O.I. system is used to analyse an observed variable without using that datum and to check the normalized departure

$$\frac{(A^o - A^i)^2}{E^{P2}} > \eta^2 \left(\frac{E^{io^2} + a^2}{E^{P2}}\right)$$

where E^{io} is the estimated error in the O.I. algorithm and a is minimum allowed value for this error. It is due to the fact that the O.I. error estimate goes to zero with infinite data coverage and cannot take into account the facts that neither the first guss nor the the structure functions (correlation functions) used in the assimalation can resolve scales smaller than several hundred km.

The ECMWF scheme uses the box method for solving the O.I system. This means that all observations from a rather large area and vertical interval are combined in one big correlation matrix M and the system is solved only once for this area by inverting M. Ideally the area should be global in order to achieve a smooth and consistent analysis everywhere. This is not computationally possible so the box method is an approximation to this. Still a large number of observations comprising several hundred data (<451 at present) are combined in the O.I. analysis and their inter-dependencies are exploited. The basic boxes are aproximate squares of sides 5.625 degrees of latitude and data selection extends from the centre out to a distance of half the diagonal plus the componenent scale length in case of data checking or twice the scale length for the final analysis. If there are more than 451 data in the box plus a minimum area extending into the surrounding boxes the base box is split into four. This can be repeated again if there are still too many data (see Fig. 15). Vertically the interval is from surface to 100 hPa or from 300 hPa to 10 hPa for a second slab.

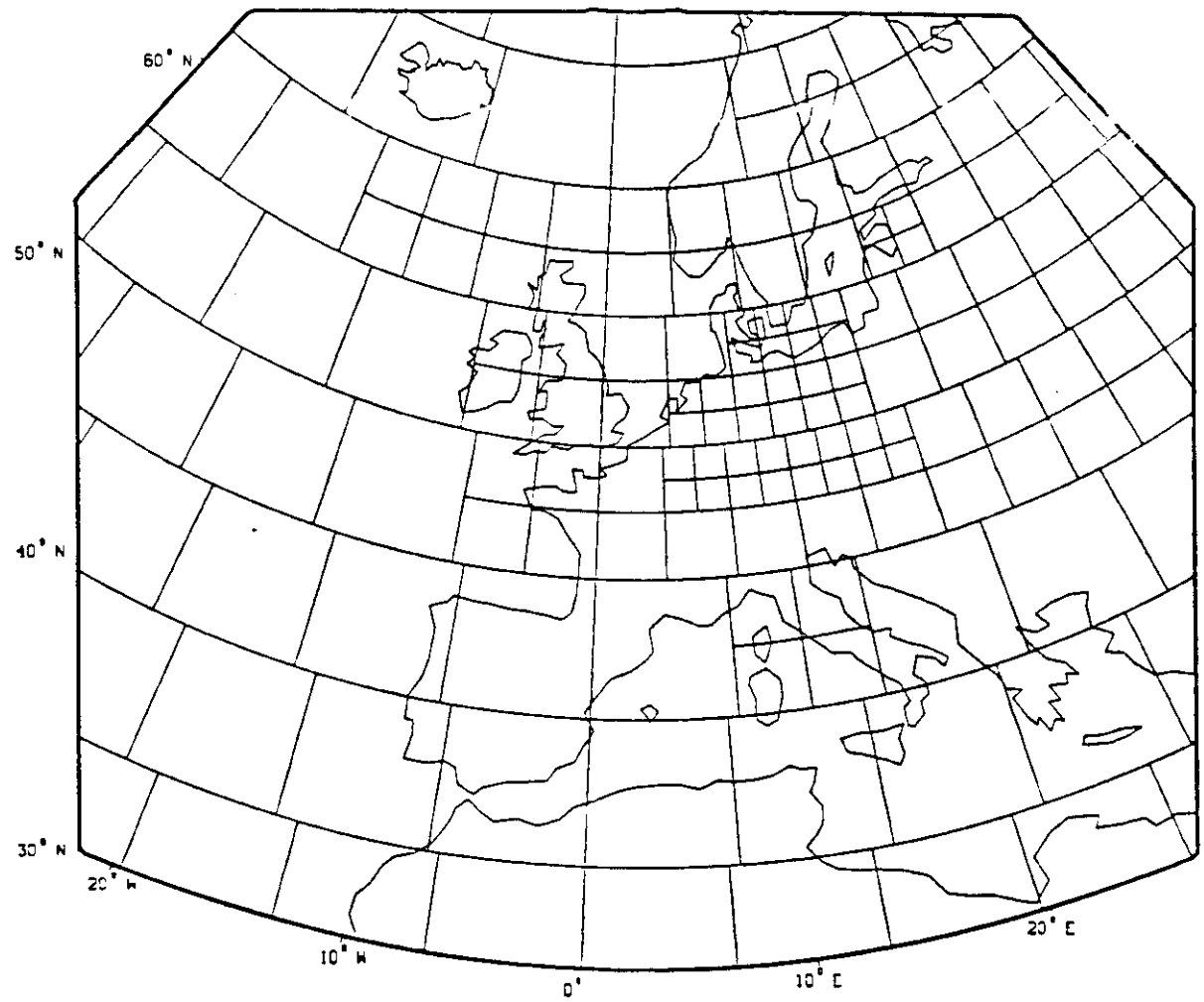


Fig. 15 Analysis box structure.

In the O.I. data checking the inverse M has to be modified so as to give zero weight to the datum to be checked.

For the gridpoint evaluation of analysis increments the matrix $\underline{\underline{M}}$ is not actually inverted since the analysis at a gridpoint can be written

$$a_k = \underline{P}_k^T \underline{\underline{M}}^{-1} \underline{B} = \underline{P}_k^T \underline{C}$$

where $\underline{B}$ is the vector of observations and $\underline{C}$ is

$$\underline{C} = \underline{\underline{M}}^{-1} \underline{B}$$

Inside one analysis box there the vector of coefficients, $\underline{C}$, is constant and independent of the gridpoint to be analysed. The same $\underline{C}$ can thus be used to evaluate all gridpoints in the box as the scalar product of the correlations $\underline{P}$ (between observations in $\underline{C}$ and the gridpoint) and the vector $\underline{C}$. $\underline{C}$ is for economy reasons computed by solving the linear system

$$\underline{\underline{M}}\,\underline{C} = \underline{B}$$

through elimination. All gridpoints inside the basic box plus a certain overlapping area outside are analysed using the same observations. Many gridpoints are analysed twice or more times due to the overlapping of analysis areas. These values are then averaged according to the proximity of the contributing boxes. Also a vertical overlap and averaging is done. This ensures that not only the analysis variables themselves but also their spatial derivatives vary smoothly.

The variables analysed are heights and winds on the model's Gaussian grid and vertical levels. Temperatures are calculated through the hydrostatic relation and the variables are spectrally transformed to the model space.

8.4 Humidity and surface analysis

The humidity analysis is a uni-variate 3-dimensional O.I system in relative humidity. Observations of temperature and dew point from SYNOP, SHIP and TEMP are converted to relative humidities. Also cloud information is used to ascribe relative humidities to certain layers. Satellite precipitable water content is converted to relative humidity and those data are also used in the analysis.

The horizontal correlation function is an exponential (Gaussian) with a length scale of 300 km.

The computation is organised in a way very similar to the mass and wind analysis. A first guess check and test for supersaturation is performed after interpolations from first guess are made and normalized departures formed. Also multi-level and O.I. checks are performed. The analysis is done using the box technique in the same way as before

with the exception that only one vertical slab up to 250 hPa is used.

Surface analyses of sea surface temperatures and snow depths are also made. For other surface variables the first guess values are kept.

The sea surface temperature (SST) data used at ECMWF are not real data but a gridded 5x5 degree analysis from NMC, Washington. The problem here is to project the SST data onto the ECMWF Gaussian grid, but only on to points which are sea points in the model, not land points or ice points. It is technically rather tricky to fit a low definition field like this onto a high resolution grid where the definitions of land, sea and ice may be imcompatible. At present (1986) this is done through a spline interpolation technique of increments on the 5x5 degree grid. It is planned to soon start using higher resolution SST and to replace the analysis method by successive correction.

Snow depth is analysed using either snow information or precipitation information from SYNOP's. A weighted average (i.e. a successive correction step with no weight to the first guess) of the reported snow depth is formed using an influence radius of 333 km.

The accumulated precipation data is converted to snow fall if the reported temperature is less than 2 C. Snow melt is estimated if the temperature at the lowest model level is above 0 C. These estimates of snow fall and snow depths are weighted together and added to the persistence (previous) analysis.

If no data are available the analysis uses a weighted average using 98 % of the first guess prediction and 2 % climatology.

REFERENCES

Bergthorsson,P. and Doos,B.R. 1955. Numerical weather map analysis. Tellus 7,329-340

Courtier,P. and Talagrand,O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. Q.J.R.Meteor.Soc. 113,1311-1328

Cressman,G.P. 1959. An operational objective analysis system. Mon.Wea.Rev. 87,367-374

Le Dimet,F.-X. and Talagrand,O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. Tellus 38A,97-110

Gandin,L.S. 1963. Objective analysis of meteorological fields. Translated from Russian by the Israeli Program for Scientific Translations. (1965)

Lorenc,A. 1979. Meteorological data analysis. Lecture Note No. 3, 68 pp. ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, England.

Lorenc,A. 1981. A global three-dimensional multivariate statistical interpolation scheme. Mon.Wea.Rev. 109,701-721

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

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