DESIGN OF ECMWF ANALYSIS SCHEME

A. Lorence

ECMWF*

* Present affiliation: UK Meteorological Office, Bracknell

1. INTRODUCTION

Design of the ECMWF analysis scheme began in earnest in 1976; by 1977 most of the major design decisions had been made, and a preliminary version was working (Lorenc et al. 1977). The first version of the current operational scheme was tested in 1978 and regular operational use began in 1979. Development of the scheme continues, and some of the features envisaged in 1977 are still not operationally implemented. The purpose of these seminars is not to attempt a full description of the operational analysis scheme, but rather to describe the various decisions made in designing the scheme and to illustrate the effect of various design features on simplified or practical examples. I shall concentrate on the mass and wind analysis.

The analysis is a global multivariate three-dimensional statistical interpolation scheme incorporated in a six-hourly intermittent data assimilation. In the following sections I shall describe and discuss each of these aspects in turn.

2. STATISTICAL ("OPTIMUM") INTERPOLATION

This technique, usually credited to Gandin (1963), combines the first-guess and any observations available with weights designed to minimize the estimated analysis error. It is thus very suitable for an analysis scheme which has to cope with differing observation types and qualities and different areas of the globe with varying first-guess accuracies. Other advantages of the technique are its potential for consistent multivariate use of different data types (e.g. height, thickness and wind) described in a later section, and the ease with which a systematic quality control of data can be included, since every value has an associated estimated error.

The derivation of the statistical interpolation equations is well known; I repeat it here for reference and to establish my notation.

The statistical techniques used are independent of the actual variables observed or interpolated, so I use a notation which does not explicitly differentiate between them, allowing subscripts to range as appropriate over all observed or analysed values whatever their position, level, or variable type. Thus B_i is any observed datum selected for the analysis, and A_k any analysed value within the analysis volume.

For all observed or analysed values I assume the existence of predicted (first-guess) values P_i, P_k and "true" values T_i T_k, the last being the value we wish to estimate in the analysis. Note that T is not necessarily the actual true value, since we do not wish to analyse atmospheric features below a certain scale.

Deviations from this "true" value are denoted by lower case letters:-

$$a = A - T \tag{1a}$$

$$b = B - T \tag{1b}$$

$$p = P - T \tag{1c}$$

All analysed, observed or predicted values have associated error estimates E defined by

$$E^a = \langle a^2 \rangle^{\frac{1}{2}} \tag{1d}$$

$$E^o = \langle b^2 \rangle^{\frac{1}{2}} \tag{1e}$$

$$E^p = \langle p^2 \rangle^{\frac{1}{2}} \tag{1f}$$

where < > indicates an average over a large ensemble of similar realizations. It is convenient to derive equations in dimensionless form, and to have symbols for deviations from the prediction, so I define

$$\alpha = a/E^a \tag{1g}$$

$$\beta = b/E^b \tag{1h}$$

$$\pi = p/E^p \tag{1i}$$

$$q = (B - P)/E^p \tag{1j}$$

$$r = (A - P)/E^p \tag{1k}$$

$$\varepsilon^o = E^o/E^p \tag{1l}$$

$$\varepsilon^a = E^a/E^p \tag{1m}$$

All the above take subscripts i (or j) ranging over all observed values, or k ranging over all analysed values, whatever their position level or variable.

The basis of the statistical interpolation method is that the analysed deviation from the prediction is given by a linear combination of N observed deviations:-

$$r_k = \sum_{i=1}^{N} w_{ik} \; q_i \tag{2}$$

with the weights (w) determined so as to minimize the estimated analysis error E_k^a.

Substituting (1) in (2) gives

$$\alpha_k \; \varepsilon_k^a = \pi_k + \sum_{i=1}^{N} w_{ik} \; (\beta_i \; \varepsilon_i^o - \pi_i) \tag{3}$$

Squaring (4) and taking the ensemble average gives

$$(\varepsilon_k^a)^2 = 1 + 2 \sum_{i=1}^{N} w_{ik} \; (<\pi_k \; \beta_i> \varepsilon_i^o - < \pi_k \; \pi_i>)$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ik} \; (<\pi_i \; \pi_j> + \varepsilon_i^o <\beta \quad \beta> \varepsilon_j^o$$

$$- \varepsilon_i^o <\beta_i \; \pi_j \quad - <\pi_i \; \beta_j \quad \varepsilon_j^o) \; w_{jk} \tag{4}$$

These summations are conveniently manipulated using a vector and matrix notation, so I define

$$\underset{\sim}{w}_k = [w_{ik}] \tag{5a}$$

$$\underset{\sim}{p}_k = [<\pi_k \; \pi_i> - <\pi_k \; \beta_i> \varepsilon_i^o] \tag{5b}$$

$$\underset{\sim}{q} = [q_i] \tag{5c}$$

$$\underset{\sim}{M} = [\, [<\pi_i \; \pi_j> + \varepsilon_i^o <\beta_i \; \beta_j> \varepsilon_j^o$$

$$- \varepsilon_i^o <\beta_i \; \pi_j> - <\pi_i \; \beta_j> \varepsilon_j^o] \,] \tag{5d}$$

(2) and (4) then become

$$r_k = \underset{\sim}{w}_k \; \underset{\sim}{q} \tag{6}$$

$$(\varepsilon^a_k)^2 = 1 - 2\, \underset{\sim}{w}^T_k\, \underset{\sim}{p}_k + \underset{\sim}{w}^T_k\, \underset{\sim}{M}\, \underset{\sim}{w}_k \qquad (7)$$

I can now proceed to the derivation of the equation for the "optimum" weights, which minimize E^a_k. Since the ensemble average < > is assumed to be over a large number of similar realizations with the same estimated errors E, this is equivalent to minimizing the normalized error variance given by (4) or (7). By equating $\partial(\varepsilon^a{}_k)^2/\partial w_{ik}$ to zero for i = 1,N we get a set of linear equations for the weights which give:

$$\underset{\sim}{w}_k = \underset{\sim}{M}^{-1}\, \underset{\sim}{p}_k \qquad (8)$$

The analysed value and estimated error corresponding to these weights are:-

$$r_k = \underset{\sim}{p}^T_k\, \underset{\sim}{M}^{-1}\, \underset{\sim}{q} \qquad (9)$$

$$(\varepsilon^a_k)^2 = 1 - \underset{\sim}{p}^T_k\, \underset{\sim}{M}^{-1}\, \underset{\sim}{p}_k \qquad (10)$$

Since $\underset{\sim}{M}^{-1}$ and $\underset{\sim}{q}$ are independent of the point being analysed it is convenient to evaluate their product once only, to give a vector of analysis coefficients $\underset{\sim}{c}$. Thus for the grid-point analysis the weights $\underset{\sim}{w}_k$ are not explicitly calculated, instead (9) becomes

$$\underset{\sim}{c} = \underset{\sim}{M}^{-1}\, \underset{\sim}{q} \qquad (11)$$

$$r_k = \underset{\sim}{c}^T\, \underset{\sim}{p}_k \qquad (12)$$

To illustrate the effect of statistical interpolation in giving significant weight to an accurate first-guess Figures 1-3 show 300mb fields from a case study which I shall be using for further examples later. Figure 1 shows the 6 hour forecast first-guess in a FGGE test data assimilation for 1200 GMT 19 January 1979. Figure 2 shows an analysis not using this first-guess, but rather climatology. Since the forecast model is quite accurate for 6 hour forecasts, while the upper air data coverage is quite sparse in the south of the region shown, the estimated error in both these 300mb height fields at 40°N 180°W is about 40 m. Figure 3 shows the result of the statistical interpolation with both information sources used. The value analysed is between that from the forecast and that from the observations alone.

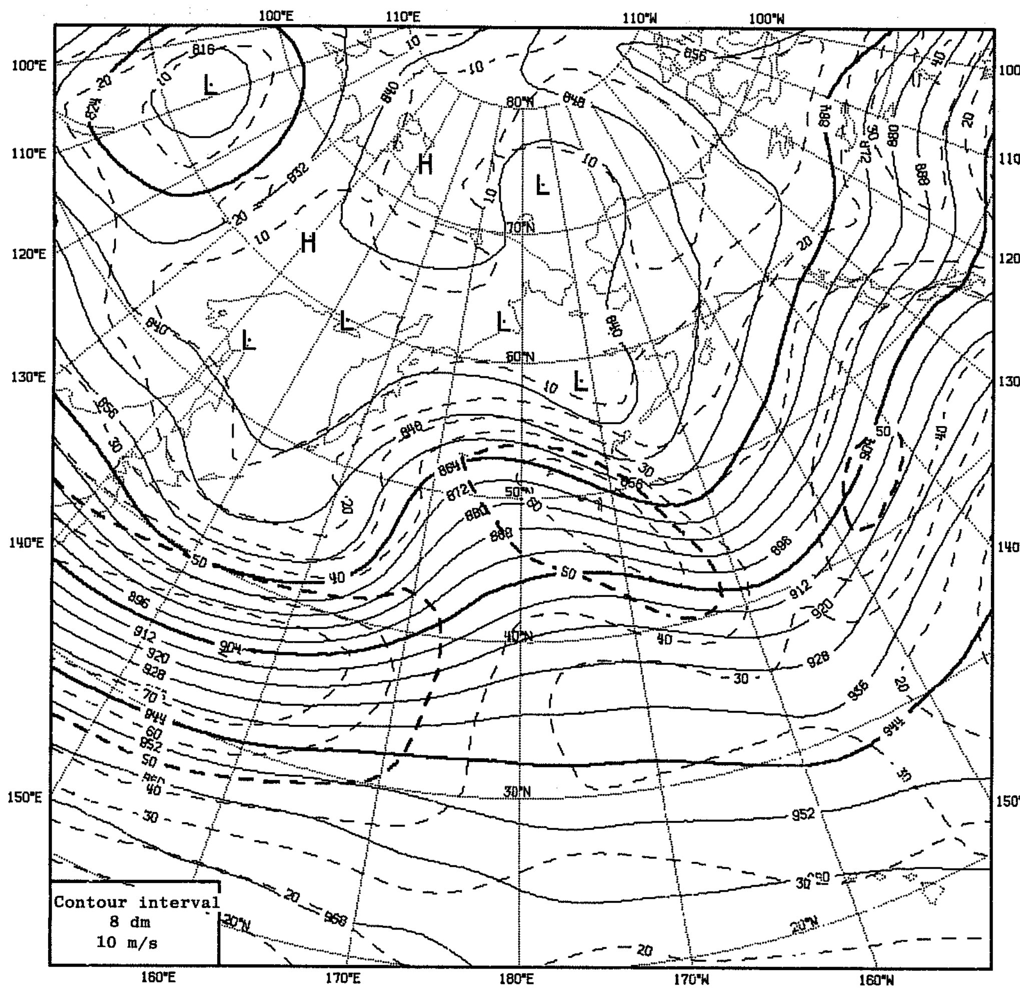

Fig. 1 300 mb geopotential height and wind speed for 1200 GMT 19 January 1979 from the 6 hour forecast used as first-guess for the analysis in the data-assimilation cycle.

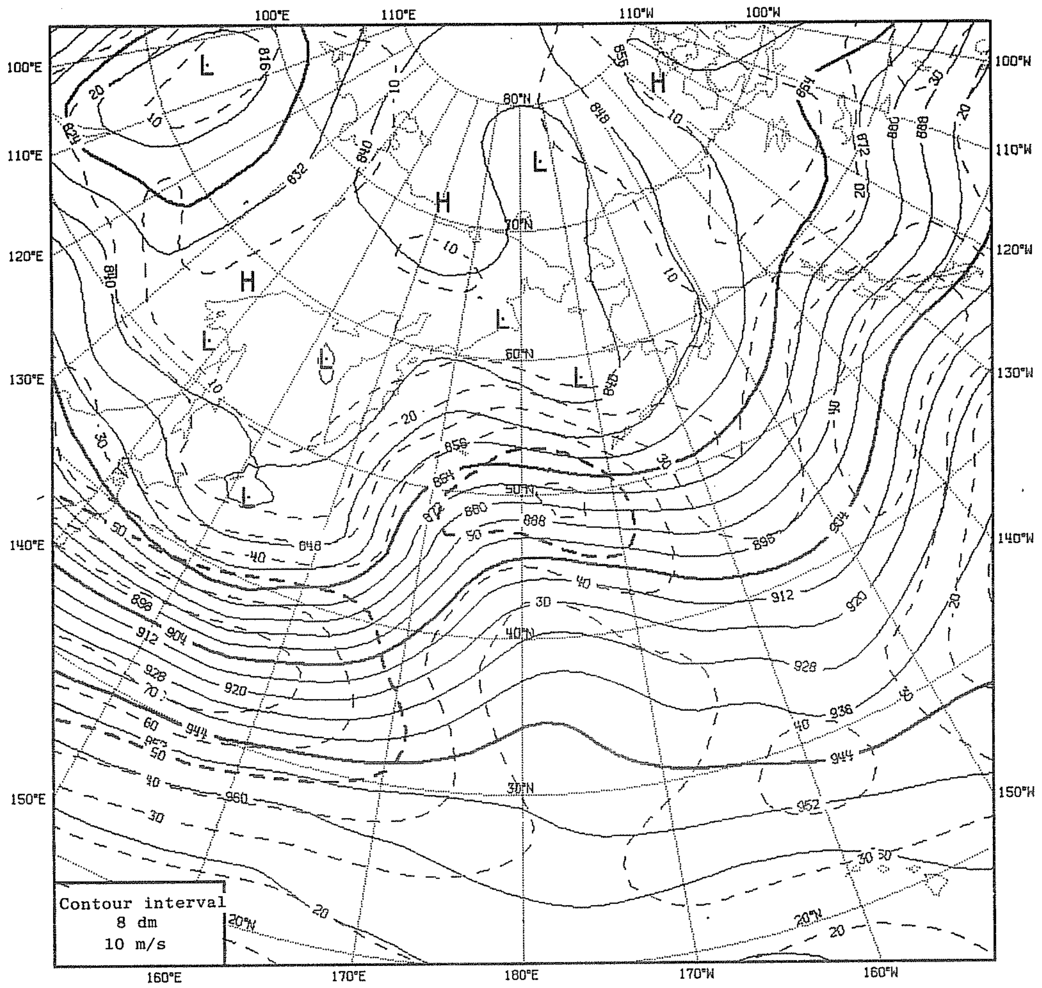

Fig. 2 As Fig. 1 for an analysis made using a climatological first-guess.

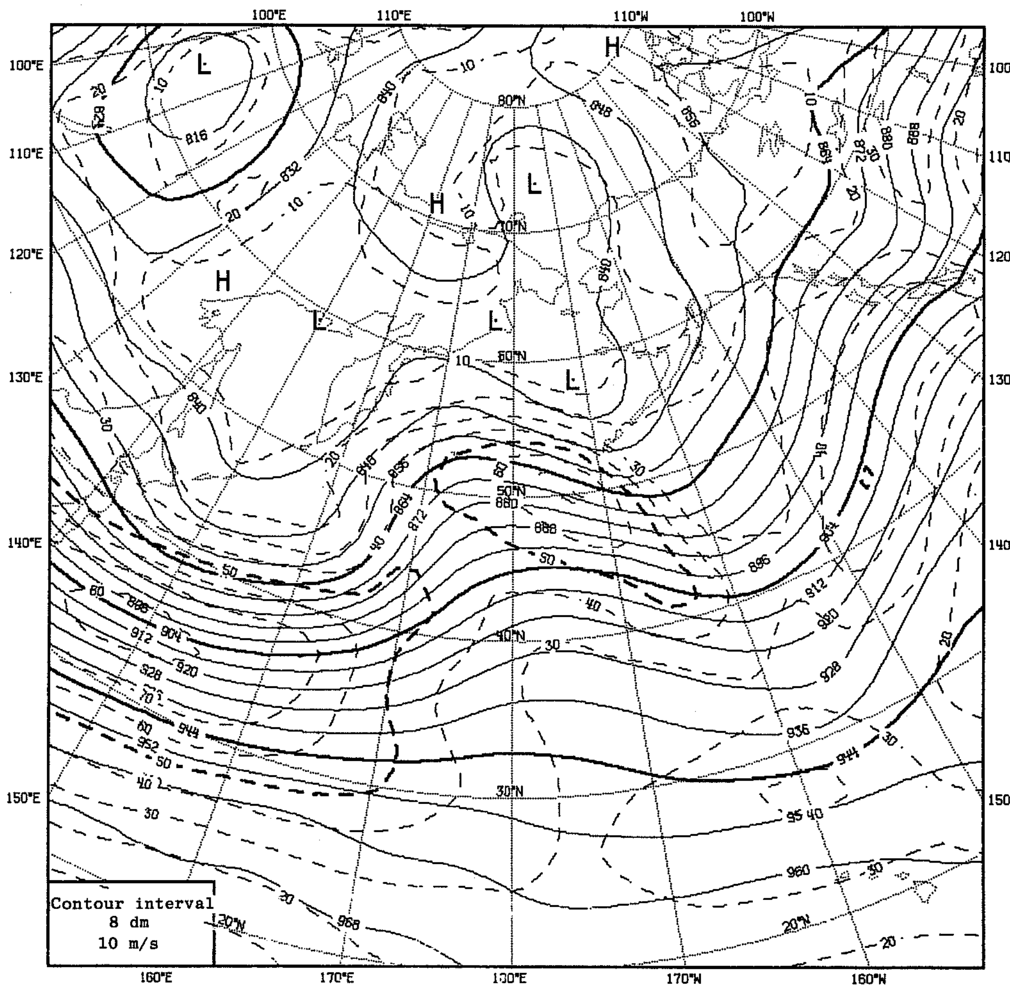

Fig. 3 As Fig. 1 for the analysis in the data assimilation cycle.

3. MULTIVARIATE ANALYSIS

A multivariate analysis scheme was considered to be desirable for ECMWF mainly because of the need to cope with "unconventional" (from a traditional European or N. American viewpoint) new observing systems which are increasing in importance, particularly in the tropics and southern hemisphere. Since conventional radiosonde observations provide data for wind height and temperature, the advantages of a multivariate scheme are marginal where these are the main source of data (e.g. Rutherford, 1973). However, to form consistent analyses using either satellite temperature soundings, or cloud motion winds and automatic aircraft reports, or ocean buoys, as the main source of data, requires some use of the multivariate relationships which the height temperature and wind fields of the atmosphere obey. So the ECMWF scheme analysis mass and wind field simultaneously, using height temperature and wind data. Three relationships are valid for the prediction error covariance model used:-

i) between geopotential heights at any two levels and the thickness between them (the thickness of course being related to the layer mean temperature);

ii) between streamfunction and wind components, implying non-divergence;

iii) between geopotential height and streamfunction, approximately equivalent to the geostrophic relationship.

Relationships ii and iii, while being useful, fairly accurate descriptions of the local behaviour of prediction errors, are not precisely true on larger scales. This is not in practice troublesome since they are only used locally in the analysis, and large scale deviations from them implied by the observations are drawn in. iii is relaxed in the tropics.

An alternative to the multivariate analysis method is to analyse each variable separately, subsequently varying the fields according to the estimated analysis errors to comply with the multivariate relationships (e.g. Jones, 1976). The present scheme was preferred to such a scheme for two reasons:-

I The automatic quality control of data which is part of the analysis is more likely to be correct if all multivariate information is available at this stage.

II The information content of satellite temperature soundings is rather complex; because of their correlated observational errors they are most accurate in providing horizontal gradients of vertical differences of the height field. The multivariate analysis can use such information effectively in conjunction with reference level observations of height or wind, as illustrated in Table 1, row e. Although in principle a variational combination of fields, taking account of their absolute and gradient errors, is possible, I know of no practical 3-dimensional multivariate scheme.

TABLE 1. Estimated analysis error (m/s) of a 500 mb wind component for various distributions of error free 1000-500 mb thickness data (t) and 1000 mb wind data (v)

	Analysis error	distance of data from analysis point (km) -500	-250	0	250	500
a	3.26					
b	2.96					t
c	2.51	t				t
d	2.37		t		t	
e	0.38		t	v	t	
f	3.17			v		

To illustrate the effect of the geostrophic relationship used to cross couple height and wind data and analyses, I use as example the surface fields corresponding to the central area of Figures 1 and 3. Figure 4 shows the 6-hour forecast first-guess with verifying observations, Figure 5 the standard analysis, and Figure 6 an analysis made not using any wind data. A comparison of the pressure fields of Figures 5 and 6 shows the effect of the cross coupling in a traditional situation of equal coverage of height and wind data. The wind data have a marginal beneficial effect on the pressure field, for instance, tightening the gradient near 43^{O}N 161^{O}E. A comparison of the wind fields of Figures 4 and 6 shows the much greater effect when data of only one type is available, as was the case in much of the southern hemisphere oceans during FGGE. The geostrophically coupled analysis made using only pressure and temperature data fits most of the observed winds quite well.

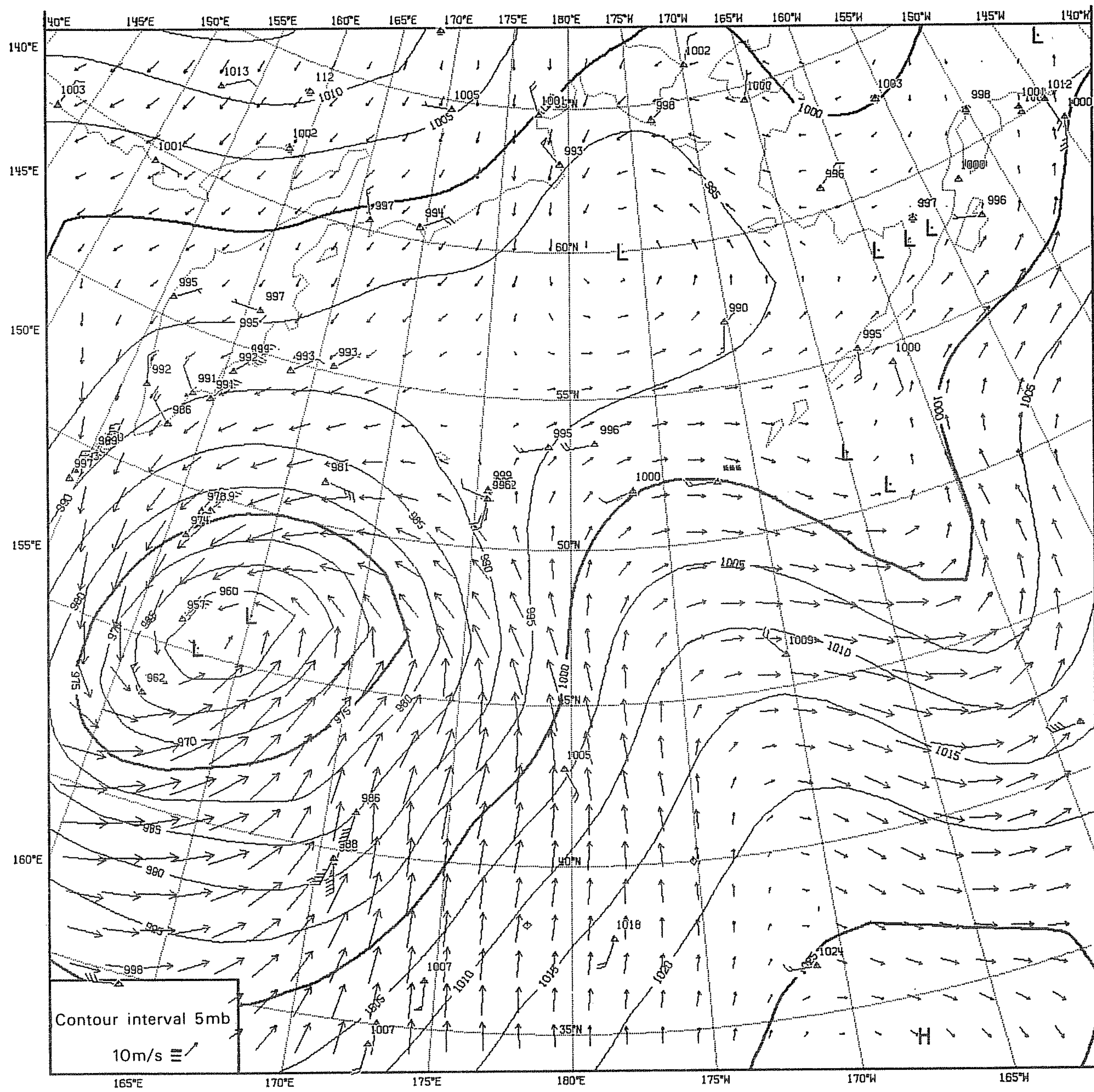

Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

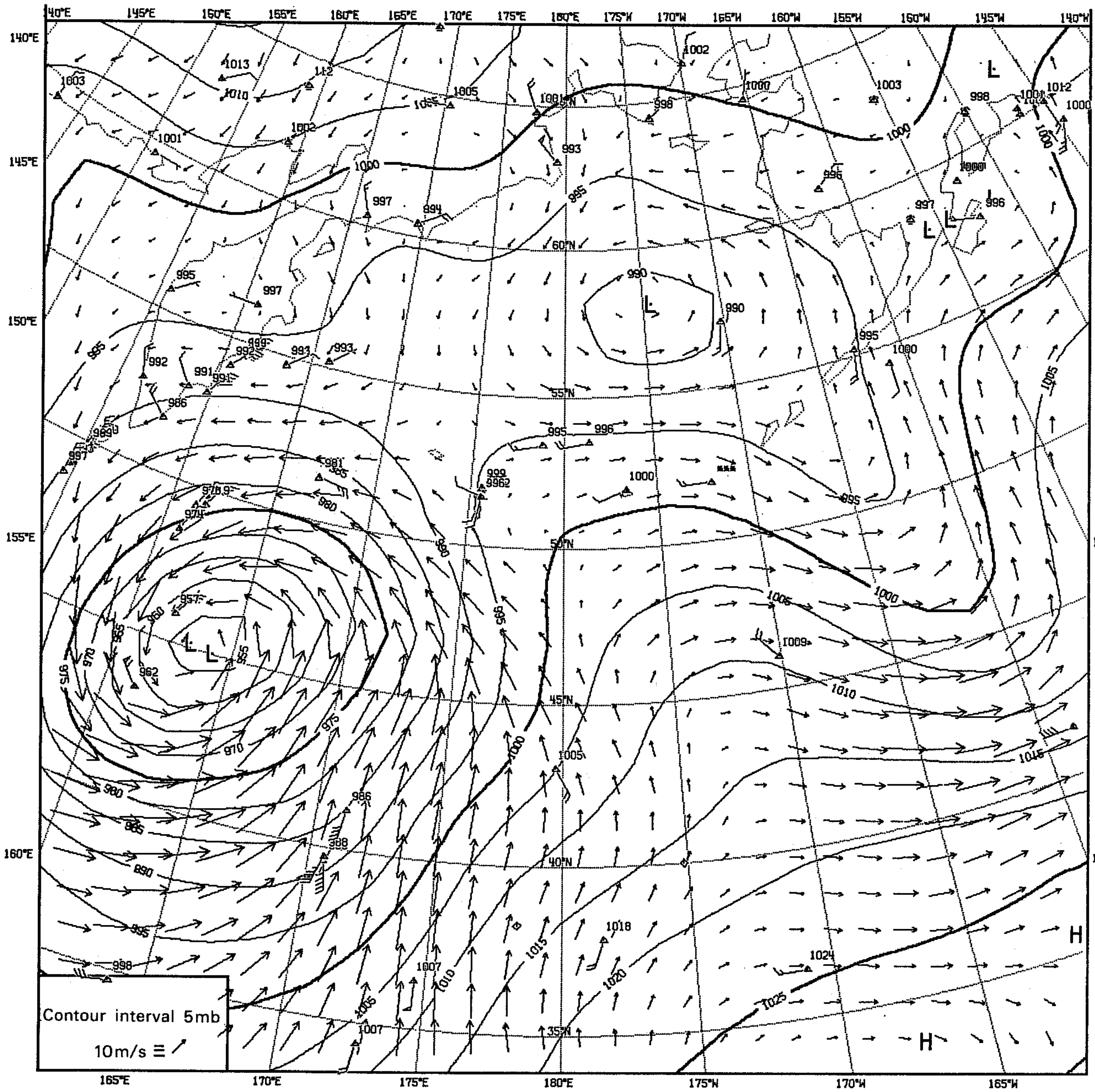

Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

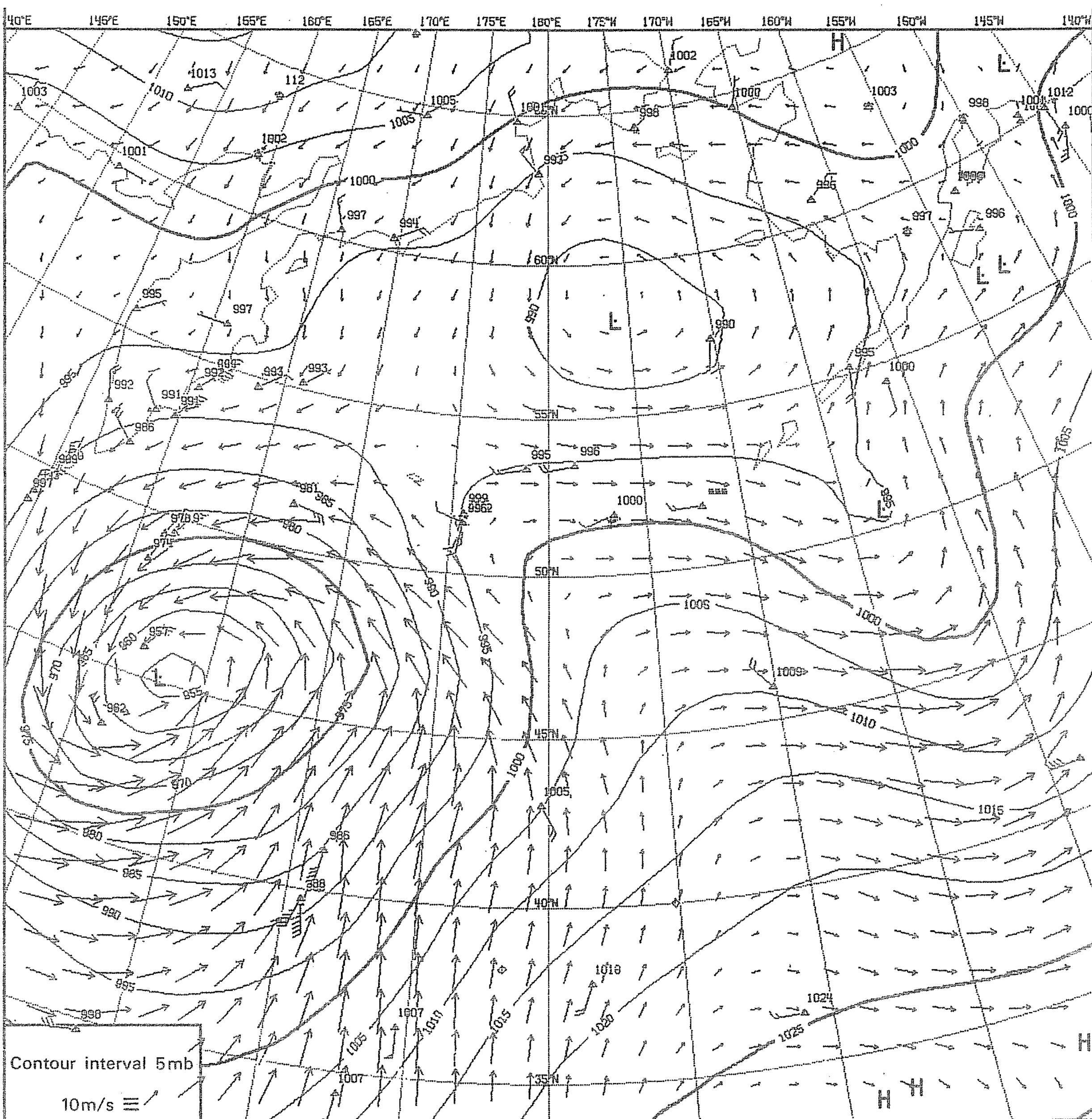

Fig. 6 As Fig. 5 not using wind data

4. THREE DIMENSIONAL USE OF DATA

The three dimensional multivariate scheme allows best use to be made of wind, height and thickness data, no matter how they are distributed; but it is computationally expensive, and cheaper schemes exist which can use most data distributions. The Canadian scheme (Rutherford, 1976) is quasi three-dimensional, interpolating data vertically and then horizontally. However, this does not allow for the optimum combined use of a thickness sounding and a nearby surface reference level pressure. The NMC scheme (Bergman, 1979) analyzes temperature and wind three-dimensionally, however, this does not allow optimum use to be made of the geostrophic relationship, which links winds and heights more directly. Schemes which analyse the levels in a pre-determined order, usually starting with the surface, and then use the analyses already made to interpolate first-guesses for the other levels and transform thickness soundings to heights (e.g. Jones, 1976), can produce three dimensionally consistent analyses provided that the data are conveniently distributed. However, if surface data are lacking while upper air data are not, as can occur over southern oceans with satellite temperature soundings and cloud motion winds, then the surface analysis needs to use these data three dimensionally. Where radiosondes, which already provide vertically consistent information at all levels, are the main data source, a two dimensional scheme is sufficient (e.g. Schlatter et al., 1975). Since a powerful computing facility was available the fully three-dimensional scheme was chosen, in order to be able to deal with all possible observing systems.

5. DATA ASSIMILATION

An analysis, if it is to be as accurate as possible, must supplement information from the currently available observations by two things:

1. Information from earlier observations.

2. Knowledge of the likely structure and scales of atmospheric motion, and of the balance which is usually observed between the various fields (mass, wind, humidity) of the atmosphere.

In a data-assimilation scheme both of these are provided by a numerical model of the atmosphere, which can update information from past observations to the current analysis, and assimilate all the data into a consistent multivariate three-dimensional analysis which represents the atmospheric motion in a realistic way. When, as at ECMWF, the main use of the analysis is to provide initial conditions for a numerical forecast, the advantage of using a numerical model for this outweighs the main disadvantage, which is that biases and inaccuracies in the model's formulation and limitations to its resolution mean that the final analysis does not always accurately represent all the detail available in the observations. Thus an accurate numerical model of the atmosphere is an integral part of our assimilation scheme. However, for flexibility in the future, we have attempted to keep the scheme as independent as possible of the details of the particular numerical model used.

Ideally the numerical prediction model should be used in such a way that each observation is inserted at the appropriate model time. However, other factors make it desirable to process batches of observations simultaneously:

1. A quality control check of each observation by comparing it with information interpolated from nearby observations is necessary before any observation is used.

2. Sophisticated analysis and initialization techniques are used to ensure that information from an incomplete coverage of observations is inserted into realistic scales of motion, with approximate balance between the various fields. These techniques require appreciable computation and so can only be justified for a large batch of data.

3. The desire to keep the scheme's organization independent of any one forecast model makes a distinct analysis - initialization-forecast cycle preferable.

Hence, we use a compromise 6 hourly intermittent data assimilation. This implies that observations up to 3 hours from the nominal analysis time are used with a forecast field valid for the analysis time. Since the major off-time observations currently available are satellite temperature soundings, whose observational errors are usually larger than atmospheric changes in 3 hours, the effect of ignoring such changes is probably negligible.

In the limit of infinitely high resolution the choice of coordinates to represent the fields would be purely a matter of convenience, since the transformation to any other is exact and reversible. For practical resolutions this is not generally the case, and the effect of such transformations can be quite large (McPherson et al., 1979). The most convenient coordinates for analysis are not in general those used in the forecast, and since we desire anyway to be as independent as possible of any particular forecast model we have accepted the need for such transformations in the ECMWF data-assimilation system. It is hoped that the relatively high resolution used, together with careful design of interpolation techniques, will minimize their effect.

The various interpolation steps are:-

1. Interpolate forecast to analysis levels and variables to give "first-guess".

2. Interpolate first-guess horizontally to observation positions.

3. Vertically interpolate (or extrapolate using first-guess gradient) observations to analysis levels and variables.

4. Subtract the results of 2 and 3 to give observed deviations.

5. Analyse (i.e. 3-dimensional multivariate statistical interpolation) the observed deviations to give analysis increments.

6a. Interpolate analysis increments vertically to forecast levels and variables, and add to forecast.

or 6b Add analysis increments to first-guess, and interpolate to forecast levels and variables.

The interpolation of increments method (6a) is more in keeping with the idea of data assimilation than the more traditional full field method (6b), since detail in the forecast which is not actually contradicted by observations is preserved. For observations of analysis levels and variables (e.g. standard level radiosonde heights and winds, SATEM thicknesses), interpolation of the analysis increments is equivalent to interpolation of the vertical covariance functions by the same method followed by a statistical interpolation direct to the forecast levels and variables. For historical reasons the full field method is still in use in the operational ECMWF scheme.

The advantages of the data-assimilation approach are most pronounced in data-space areas and for parameters sensitive to the internal consistency of the analysis, such as the vertical velocity. Figure 7 shows the vertical velocity at 700 mb corresponding to Figure 5. Note the vertical motions consistent with a typical warm sector depression near the main low. Without a numerical model forecast as first-guess, (or the skill of an experienced meteorologist in using the warm sector depression concept), it is impossible to analyse such patterns from the available data, although some gross features may be obtained. The vertical motions in the cast of Figure 7 are associated with the diffluent upper trough (Figure 3) which is very poorly defined by the available data alone (Figure 2). Thus they are absent in the analysis from climatology as a first-guess.

The disadvantage of the data-assimilation method is that the model limits the resolution available. This is further limited in the ECMWF scheme by the different vertical grid used for the analysis computations. Information from the observations is truncated to this latter resolution, as is the information from the forecast when the full fields are interpolated from the analysis grid to the forecast grid. When only the analysis increments are interpolated forecast information is not affected by the analysis resolution, as illustrated in Figure 8. Figures 8 a, b, c, d show respectively the forecast temperatures, the first-guess temperatures in analysis coordinates (layer mean values assigned to the middle of each layer), the analysed temperatures, and the corrected forecast model temperatures after the addition of the interpolated analysis increments. On each is also plotted the temperatures from the nearest observation. The forecast (Fig.8a) is too warm in the

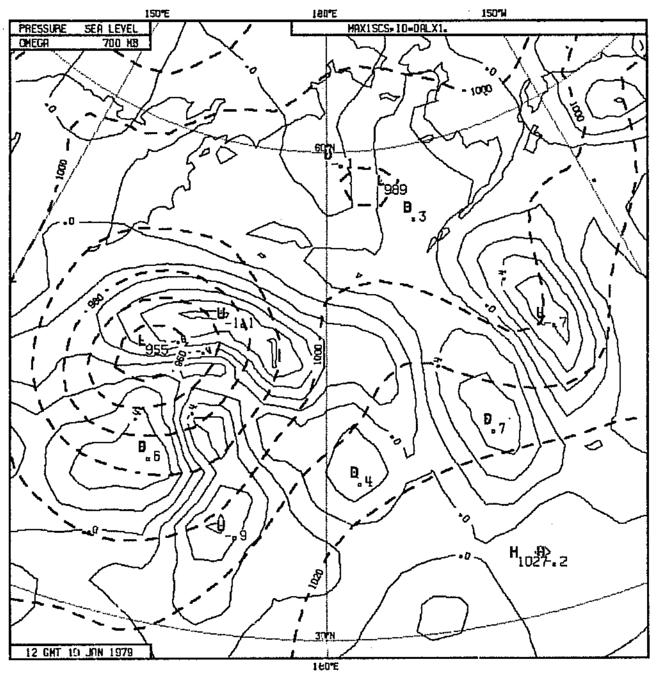

Fig. 7 **700 mb vertical velocity analysis (pascals/sec) corresponding to Fig. 5, with sea level pressure shown dashed.**

lower troposphere, consistent with the low and associated warm front being too advanced. The first-guess in analysis coordinates (Fig. 8b) reflects this too; note that the details of the boundary layer have been lost. Since the height data from the observation, not the temperatures, are used in the analysis only the average temperatures of the analysis are correct (Fig. 8c). Finally, the forecast coordinate analysis (Fig. 8d) similarly has reasonable average temperatures, and also has retained the realistic boundary layer structure of the original forecast.

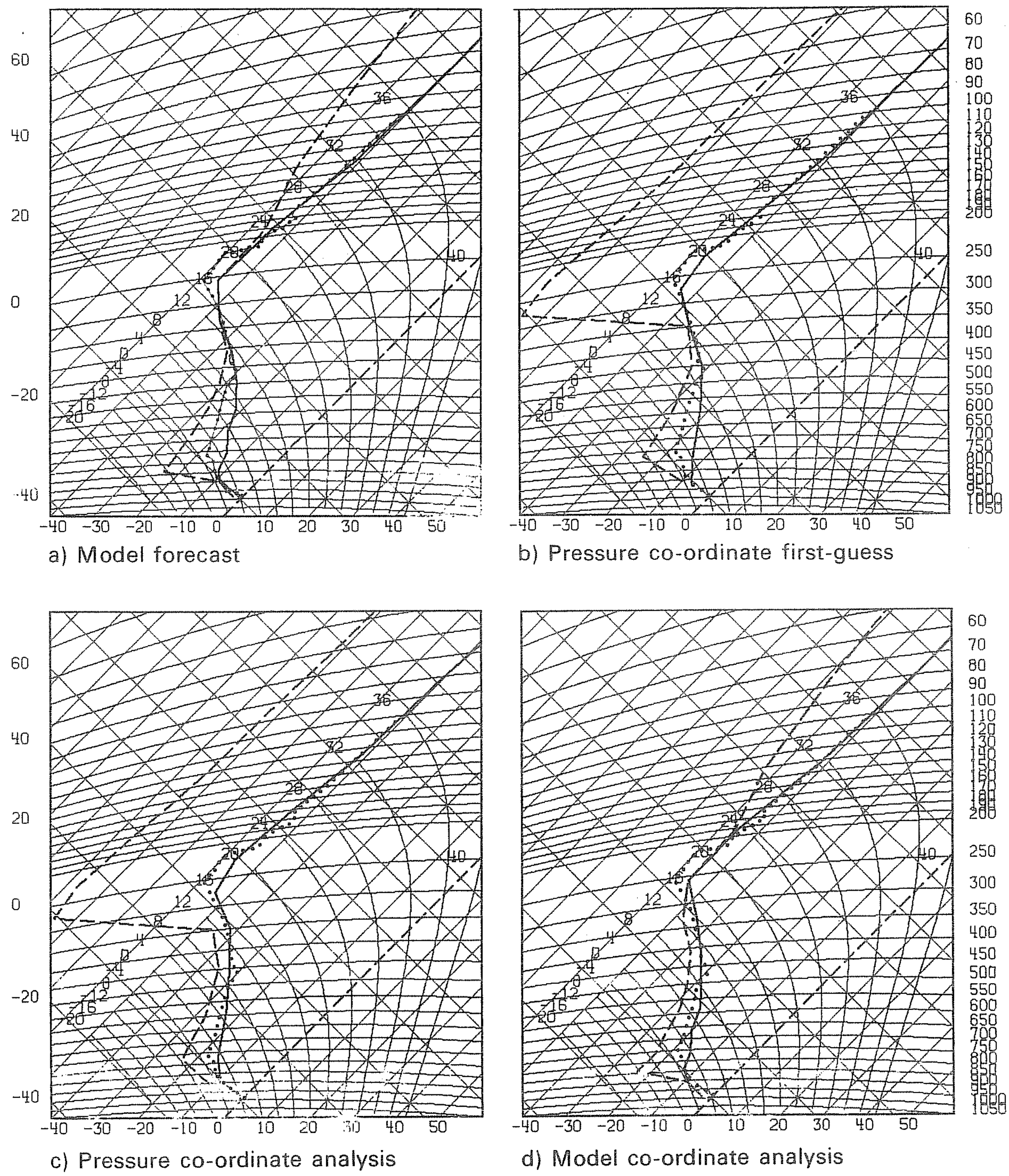

Fig. 8 Temperature soundings for the grid point at 52.5°N 174.4°E, 12 GMT 19 January 1979 at various stages during the interpolation/analysis process, with the nearest observation (at 52.7°N 174.1°E) shown on each for comparison.
______temperature ----- dewpoint ....observed temperature.

6. DATA SELECTION AND ORGANIZATION

A novel feature of the ECMWF analysis scheme is the organization of the data selection and solution of the statistical interpolation equations. The scheme was designed for a vector processing computer especially suitable for the efficient solution of large linear systems of equations. On the other hand the logical operations required for selecting only the best data in order to keep the systems small do not exploit the computer's full speed. Indeed the design of algorithms to decide which data are best in a three-dimensional multivariate analysis is extremely difficult. So instead of carefully selecting a few data, typically 8 to 15 values, as in other analysis schemes (e.g. Rutherford (1976), Schlatter (1975), Bergman (1979)), the ECMWF scheme uses most nearby data, usually between 100 and 200 values.

This enables the full potential of the analysis method to be exploited, since within such a large number of data it is possible to include height wind and thickness data for several layers of the atmosphere. It is neither necessary nor practicable to set up and solve the large systems of equations this entails for each analysis grid point and variable. Instead this is done for analysis volumes about 660 km square and, in data rich areas, a third of the atmosphere deep. In data sparse areas the full depth of the atmosphere is done in one volume. The same selection of data and solution of the equations is used to check all the data within a volume and to evaluate the analysis for height and wind at all analysis grid points in it. In fact, in order to avoid discontinuities at analysis volume boundaries (particularly important for derived quantities such as divergence) the analysis is evaluated also for surrounding grid points as far as the centre of neighbouring analyses volumes. Several analysed values are thus obtained for each point which are then averaged with weights varying linearly from 1 at the volume centre to 0 at a neighbouring volume centre.

7. COVARIANCE MODEL

A feature of the statistical interpolation method, best seen when the equations are expressed as in (12), is that the analysed increment field is a linear combination of prediction error covariance functions (here I use "field" in a multivariate three-dimensional sense). For traditional schemes, where the equations are set up and solved for each grid value, this has little significance since only one value from each functional surface is used. But for the ECMWF scheme (12) is used to evaluate the analysis over a large region of space and for both height and wind, so the properties of the prediction error covariance functions are critical. Any linear property of the functions will also hold for the analysis increments within the range of validity of one set of coefficients $\underset{\sim}{c}$. Thus deviations from such a property evident in the data can only be drawn to in so far as the coefficients change from one analysis volume to the next. For instance the covariance functions used for wind components imply non-divergence, so any divergences in the analysed increments occur uniformly over the overlap region of analysis volumes (Described at the end of the last section), with an effective resolution of only 660 km.

The larger data coverage and region of validity of each set of analysis coefficients makes the use of a correct scale for the horizontal structure function of the prediction error covariances more critical. Figure 9 shows a simple one-dimensional analysis of idealized surface pressure data representing a 1000 km wave, with various values of the horizontal scale parameter s (the standard deviation of the Gaussian structure function). For more traditional schemes which only select a few data for each grid point, the sensitivity to s is much less (Figure 10).

Because of this sensitivity to the covariance model, this is probably the area where future research might enhance the scheme. Possible avenues are:

1. Automatic recognition of certain meteorological situations, and the specification of appropriate error covariances.

2. Specification of velocity potential error covariances in addition to the current height and streamfunction, together with appropriate cross covariances with height and streamfunction, to analyse details in the divergent wind.

3. Relaxation of the vertical x horizontal separability and horizontal isotropic assumptions, to allow specification of tilts for developing systems (Bengtsson 1980) and non-isotropic frontal structures.

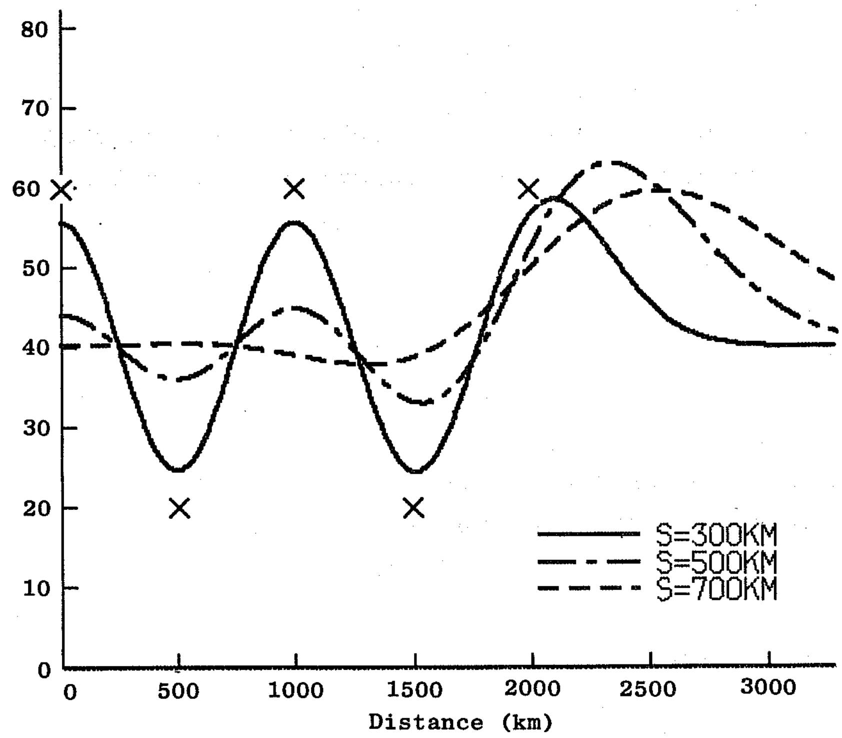

Fig. 9 Analyses of a 1000 km wave from 9 observations marked x for various horizontal prediction error correlation scales (s). Only half of the symmetric situation is shown.

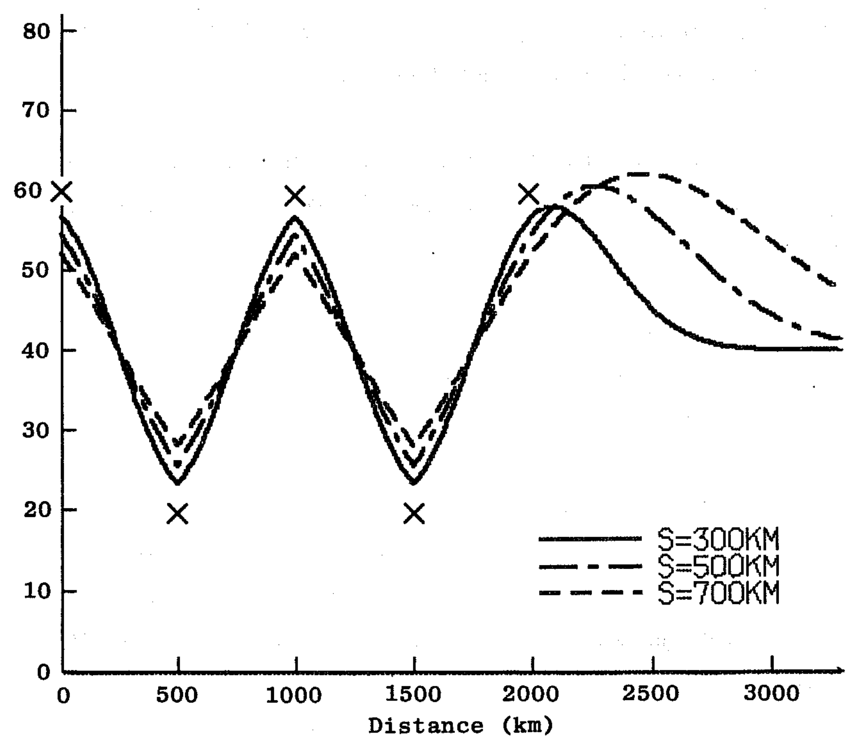

Fig. 10 As Fig. 9 selecting only the 2 nearest observations to each point.

References

Bengtsson, L. 1980 On the use of a time sequence of surface pressures in four dimensional data assimilation. Tellus. 32, 189-197.

Bergman, K.H. 1979 Multivariate analysis of temperatures, and winds using optimum interpolation. Mon. Wea. Rev. 107, 1423-1444.

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

Jones, D.E. 1976 The UK Meteorological Office objective analysis scheme for GATE. Proceedings of the JOC study group conference on 4-dimensional data assimilation. GARP program on numerical experimentation Report No. 11. 122-131.

Lorenc, A., Rutherford, I. and Larsen, G. 1977 The ECMWF analysis and data assimilation scheme - Analysis of mass and wind fields. ECMWF Tech. Report No. 6, ECMWF, Shinfield Park, Reading, England, 47 pp.

McPherson, R.D., Bergman, K.H., Kistler, R.E., Rasch, G.E. and Gordon, D.S. 1979 The NMC operational global data assimilation scheme. Mon. Wea. Rev., 107, 1445-1461.

Rutherford, I. 1973 Experiments on the updating of PE forecasts with real wind and geopotential data. 3rd cont. on Prob. & Stats. in Atmos. Sci. Preprints, American Meteorological Society.

Rutherford, I. 1976 An operational 3-dimensional multi variate statistical objective analysis scheme. Proceedings of the JOC Study Group conference on 4-dimensional data assimilation (Paris, Nov. 1975). The GARP programme on numerical experimentation. Report No. 11. WMO/JCSU 98-111.

Schlatter, T.W., Branstator, G.W. and Thiel, L.G. 1976 Testing a global multivariate statistical objective analysis scheme with observed data. Mon. Wea. Rev. 104.