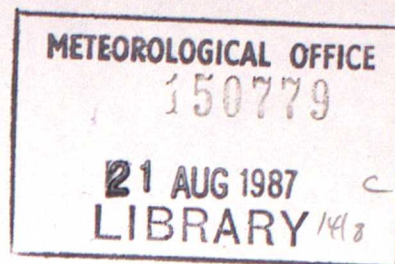


287

LF/MO11



MET O 11 TECHNICAL NOTE 224

FOUR-DIMENSIONAL ANALYSIS BY REPEATED INSERTION OF OBSERVATIONS INTO A NWP
MODEL

by

A C Lorenc and R Dumelow

December 1985

revised July 1987

Met O 11 (Forecasting Research)

Meteorological office

London Road

Bracknell

Berkshire RG12 2SZ

This paper has not been published. Permission to quote from it should be obtained from the Assistant Director of the above Meteorological Office branch.

1. INTRODUCTION

The four-dimensional analysis problem for numerical weather prediction (NWP) is to deduce the 'best' initial conditions for a NWP forecast model on the basis of available observations and prior knowledge about atmospheric behaviour. One type of prior knowledge is that of equations governing the evolution of atmospheric fields, as encapsulated in a NWP forecast model. The use of this together with a four-dimensional distribution of observations is known as four-dimensional data assimilation.

One technique of four-dimensional data assimilation is that of repeated insertion of observed values into a forward running NWP model. This has been developed for research and later operational use at the Met Office over a number of years (see Lorenc 1984a for a review). Note that the method involves repeated insertion of observed values, rather than of analysed fields, and thus has some similarity to the successive correction analysis method. In this it differs from several other repeated insertion schemes (eg Stern et al 1984).

In this note we describe the latest version of this technique, developed for use by the operational forecast models at the Met Office (a global model with 150 km resolution and a limited area model covering the North Atlantic and Europe with 75 km resolution). In section 2 the method is described in terms of the theory developed in Lorenc (1986). In section 3 we give a more practical description of the implementation of the method

as an efficient analysis algorithm on a vector processing computer. In section 4 we show an example analysis produced by the scheme, and in section 5 we discuss future developments.

2. THEORETICAL DESCRIPTION

a. Analysis as an inverse problem

This description uses the notation of Lorenc (1986), which should be referred to for a more complete discussion of the theory, derivation of the equations, and comparison with other analysis methods.

The NWP model, for which we require an analysis, represents the atmosphere at any one time by a vector of coefficients \underline{x} . The observed values are represented by a vector \underline{y}_0 . Given \underline{x} , we can estimate values for all the observed parameters using

$$\underline{y} = K_n(\underline{x}) \quad (1)$$

This forward process represented by K_n is often just a simple interpolation, but it can be generalized to include extrapolation in space and time. We wish to solve the inverse problem of estimating \underline{x} given \underline{y}_0 . This problem is underdetermined, for instance for the current global model \underline{x} has about 1.4 million elements, while we use only about $1/4$ million data per day. We can proceed by providing prior information about the desired solution, from a forecast from

earlier data and from linearizable constraints such as geostrophy. it can then be expressed as a variational problem; finding the x which minimizes

$$J = (\underline{y}_O - K_n(\underline{x}))^T (\underline{Q} + \underline{F})^{-1} (\underline{y}_O - K_n(\underline{x})) + (\underline{x} - \underline{x}_b)^T \underline{B}^{-1} (\underline{x} - \underline{x}_b) \quad (2)$$

where \underline{Q} is the error covariance of the observations, \underline{F} measures their representativeness, and \underline{B} is the prior background error covariance

$$\underline{Q} = \langle (\underline{y}_O - \underline{y}_t)(\underline{y}_O - \underline{y}_t)^T \rangle \quad (3)$$

$$\underline{F} = \langle (\underline{y}_t - K_n(\underline{x}_t))(\underline{y}_t - K_n(\underline{x}_t))^T \rangle \quad (4)$$

$$\underline{B} = \langle (\underline{x}_b - \underline{x}_t)(\underline{x}_b - \underline{x}_t)^T \rangle \quad (5)$$

We use subscript t to indicate the true value. If $\underline{x}[u]$ is an estimate of the minimizer of J , and if K_n can be linearized so that

$$K_n(\underline{x}[u] + \delta \underline{x}) = K_n(\underline{x}[u]) + \underline{K} \delta \underline{x} \rightarrow \quad (6)$$

then a better estimate is given by

$$\underline{x}[u+1] = \underline{x}[u] + \underline{Q}(\underline{W}(\underline{y}_O - K_n(\underline{x}[u])) + \underline{x}_b - \underline{x}[u]) \quad (7)$$

\underline{W} is a matrix of weights for each observation at each grid point, \underline{Q} is a normalization factor which controls convergence. The optimal forms are

$$\underline{W} = \underline{B} \underline{K}^T (\underline{Q} + \underline{F})^{-1} \quad (8)$$

$$\underline{Q} = (\underline{W} \underline{K} + \underline{I})^{-1} \quad (9)$$

b. The successive correction approximation

These ideal equations need approximating for practical solution; we now proceed to describe a set of approximations which constitute the successive correction method. The background error covariances are modelled as the product of a continuous correlation function μ and a variance field. The latter is assumed to be slowly varying, so it

can be interpolated to the observation positions and combined with the observational errors, which are assumed to be uncorrelated so that $O+F$ is diagonal. With multiplication by \underline{K} replaced by direct evaluation of μ at observation positions, (8) becomes

$$w_{ji} = \mu_{ji} \frac{b_{ii}^2}{o_{ii}^2 + f_{ii}^2} \quad (10)$$

where lower case letters denote elements of the corresponding upper case matrices. A good approximation for the normalization factor Q is less easy to find. If μ is a simple local correlation, always positive, and if the observation distribution is uniform, then $\underline{WK} + \underline{I}$ is nearly diagonal, and Q can be approximated by the inverse of a similar diagonal matrix. We therefore replace μ by a simple function, which is always positive and which varies more smoothly, and approximate (9) by

$$q_{jj} = (\sum_i w_{ji}' + 1)^{-1} \quad (11)$$

w_{ji}' is the weight calculated using the modified μ , and can be regarded as the contribution of the i th observation to the local observation density at j .

This approximation for Q removes much of the interaction between nearby observations implicit in (9). We rely on iteration to restore these effects (eg extrapolation of observed gradients). The forcing towards the background provided by $\underline{x}_b - \underline{x}[u]$ in (7) can be thought of as preventing the limit as u tends to infinity from fitting the

observations exactly, and 'forgetting' the original background. We start from $x[0]=x_b$, and perform only a limited number of iterations, so this term is omitted, giving

$$x[u+1]=x[u]+QW(y_o-K_n(x[u])) \quad (12)$$

This derivation leads to an analysis method very similar to that derived more pragmatically by Bergthorsson and Doos (1955) and Cressman (1959). It has the advantage of indicating how to allow for observations with differing observational errors (through $O+F$), and how the method relates to other methods such as optimal interpolation (OI), which proceed by reorganizing the matrix manipulations of (7), (8) and (9), and severely limiting the number of observations processed simultaneously.

c. Iteration steps, and constraints

Since we are relying on interaction via the successive estimates $x[u]$ to restore interactions between observations, lost when approximating (9), it is advantageous to subdivide each iteration into steps processing different types of data, and to let these also interact via modified estimates of $x[u]$. Such a subdivision also enables prior information and constraints between different model variables to be used without having to apply them in a simultaneous multivariate analysis step.

Each iteration is split into steps processing different types of data, and is alternated with a time step of the assimilation forecast model. The interpolation K_n is generalized to include extrapolation in time, so that observations valid for a period surrounding the current assimilation model time are included in each iteration. At present simple persistence is used for this time extrapolation. The actual time of each observation is used, thus at any iteration some observations might be near the end of their iterative insertion period, while later observations are just starting.

For each step, analysing only one type of data, constraints on the analysis increments are imposed by analysing increments corresponding to only a subset of the model's parameters, calculating the rest through the application of the constraint equation (eg geostrophy). Constraints on smoothness can also be imposed if desired by filtering of the increment field, although this should not be necessary if an appropriate correlation function μ is used.

d. Vertical-Horizontal splitting of analysis steps

Rutherford (1976) proposed a scheme, couched in terms of the reorganized (OI) versions of (7), (8) and (9), for splitting a three-dimensional analysis into successive vertical and horizontal steps. The splitting depends on the background error correlation function being expressible as the product of a vertical factor independent of level. As proposed the method in fact approximates the interaction between data at different horizontal positions and

different levels. A similar splitting is thus appropriate for our scheme, since such interactions are anyway ignored in the successive correction approximation, and the splitting can be regarded as a simple rearrangement of the calculation to take advantage of the factorizability of the correlation function. The OI viewpoint is however useful to us, since in the split scheme's vertical step the order of the equations to be solved is low, and the full successive correction approximations are unnecessary. We can thus, if the characteristics of the observation warrant it, perform the vertical step using more complete equations which take account of vertical correlations of the observation error, and of differences between the parameters observed and model variables.

In the split OI scheme the results of a vertical analysis at each observation or sounding position are used as 'observations' for separate two-dimensional analysis at each level. In doing this care must be taken that background information is not used twice, and hence given too much weight. One way of doing this is to constrain the vertical OI analysis such that the analysed value at each level contains no information from (ie is uncorrelated with) the background at the same level. Such analysed values can then be treated as ordinary data in the horizontal analysis at that level. Equations for this are derived in Lorenc (1981, section 3b).

3. PRACTICAL IMPLEMENTATION

a. Steps for each type of data

The model's basis is split into surface pressure, potential temperature, wind, and relative humidity, and observations are split into groups affecting these categories. The steps within each iteration are:

- i. Analyse surface pressure.
- ii. Adjust potential temperatures in accordance with a constraint that the geopotential height near tropopause level should not change.
- iii. Analyse potential temperature.
- iv. Calculate the wind increments in geostrophic balance with the changes made in steps i, ii, iii. A fraction of these increments (zero at the equator) are added to the model's winds.
- v. Analyse winds as a vector field, with correlations between components consistent with a constraint that wind increments are approximately horizontally non-divergent.

vi. The pressure and temperature increment required to balance the geostrophic part of these wind increments could be calculated and added. This possibility has yet to be tested.

vii. Analyse relative humidity.

viii. Advance the assimilation model one time-step.

b. Splitting into vertical and horizontal analysis

For the three-dimensional analyses in steps iii, v and vi above, the correlation function μ is split into the product of a vertical factor independent of position, and a horizontal factor independent of level. This enables the horizontal factor to be separated in the evaluation of $W(y_0 - K_n(x))$ in (12) and of $\Sigma w_{ji}'$ in (10), giving large savings in the amount of computation needed to process vertical soundings. Thus the stages in the processing of soundings and of single level upper-air observations, are:

i. Horizontal interpolation of the relevant model variables and levels to the observation position.

ii. Vertical interpolation, and if necessary transformation of variables, to the levels and variables observed, giving $K_n(x)$.

iii. Subtraction, to give observation deviation from background $y_0 - K_n(x)$.

iv. Vertical analysis of these, taking into account the vertical factor of μ as well as any vertical correlations in the observation errors, to give a vertical column of increments for the model's variables and levels, and also estimates of errors in these.

v. A horizontal analysis at each model level, using (12). The vertically analysed increments are used at each level as data, and their errors used in (10) to calculate the weights. The horizontal correlation factor μ in (10) is the same in the analysis for each level.

c. Correlation function

In our iterative analysis, as in the successive correction method, it is possible to compensate to a certain extent for deficiencies and uncertainties in the correlation function by changing (usually reducing) its horizontal scale each iteration. Such a change also helps speed the convergence of the iterative process (Barnes 1973). Thus only a simple approximation to the 'best' form (which we do not know anyway) should be sufficient. The form chosen has been suggested various times in the literature (eg Balgovind et al 1983, Thiebaut 1985). It has a better resolution of small scale details, and a better representation of the structure of non-divergent wind errors than the 'Gaussian' ($\exp(-1/2(r/s)^2)$) used in many optimal interpolation schemes.

For pressure, temperature and humidity we use

$$\mu_{ji} = (1 + r/s) \exp (-r/s) \quad (13)$$

where r is the distance from observation i to grid point j and s is the correlation scale. For wind, we assume that (13) holds for streamfunction, and assume that wind errors are non-divergent. This gives:

$$\mu_{LL} = \exp (-r/s) \quad (14)$$

$$\mu_{TT} = \mu_{LL} - (r/s)\mu_{LL} \quad (15)$$

where μ_{LL} is the correlation between wind components along the line from i to j , and μ_{TT} is the correlation between wind components transverse to this.

d. Vectorization

Modern vector processing computers only work efficiently if they are doing a large number of similar operations at one time. The traditional way of organising analysis schemes, grid point by grid point, is not well suited to this. Observations of differing types are randomly distributed round each point, and the processes of searching for these, and accounting for their various characteristics, are difficult to vectorize.

This scheme is structured around observations rather than grid points, so that all observations of one type can be processed together. Thus any special processing which depends on observation type, eg in 3.b.iv, is readily vectorized. This organisation also eliminates all searching, since the analysis grid is known, and it is possible to calculate directly which grid points are near each observation. It does however require more computer memory, since increments and normalization factors are accumulated simultaneously for the entire field being analysed.

e. Interaction with forecast model

The alternation of iterations of the analysis equation with steps of the forecast model has some important consequences, which give the method an advantage over a similar three-dimensional successive correction analysis. The time extrapolation in K_n is a trivial persistence forecast, and modes of the model forecast which change relative to this during the insertion period will tend to be damped by repeated insertion of constant observation values, and stationary modes excited instead. If we define 'balanced modes' as those which vary slowly in comparison to the insertion period, the method thus tends towards a full nonlinear balance consistent with the observations and the forecast model's dynamics.

There is also a practical advantage in that asynoptic observations can be used at their valid time with little extra complexity.

4. An example analysis

This example is taken from the first trial of the scheme which occurred between 2-15 June 1986. During this period the analysis correction (AC) scheme was run continuously, and one 72 hour forecast was run each day from 12Z data. Since only those observations with validity time around the time (T) of the end of the assimilation period were used, full weight was not given during the trial to those observations valid at beyond the nominal analysis time, for example, those at T+2 or T+3 hours. This, in particular, means that certain aircraft reports (AIREP) were not taken full account of during the experiment.

The example shown is from the main run analysis for 12Z 8/6/86. In figures 1, 2, 3 (where a is the operational run and b the trial) the northern Atlantic mean sea level pressure, 500 mb height with 1000-500 mb thickness and maximum winds are shown. It can be seen from all charts that the features represented are very similar but the charts from the trial are much smoother. This is particularly evident in the maximum wind charts, but the smoothing effect seems to have resulted in a reduction of the jet strengths by as much as 10-15 knots. However, the 24 hour forecast from this analysis was judged by CFO to have produced maximum winds which were

very similar to the operational forecast. Apart from the effect of smoothing, another reason for the lack of strength in the T+0 jets may have been due to insufficient weight being given to off-time aircraft reports.

Figures 4 and 5 compare some southern hemisphere analyses from the AC and operational schemes. Here there is a significant difference between the analyses with the AC scheme analysing a surface low to the west of Chile (which is reflected in the 1000-500 mb thickness) which is almost completely absent in the operational scheme. This leads (Macpherson, private communication) to a dramatic difference in the 72 hour forecasts (Figure 6) with the AC scheme producing a significant low to the east of Uruguay which is completely absent from the operational run. The AC scheme compares well with observations valid at that time, which suggested a low with the same central pressure and in a similar position as the one from the AC forecast.

5. FUTURE DEVELOPMENTS

a. Tuning

The understanding given by the theoretical derivation in section 2, rather than the more pragmatic development of the successive correction method, helps in specifying reasonable values for parameters in the method. However empirical tuning and testing is still needed for things like the insertion period and the correlation scale.

b. Differing observation characteristics

The processing of observations in batches of the same type, especially in the vertical interpolations of 3.b.ii and iv, enables account to be taken of the individual characteristics of different observing systems, while retaining vectorized code. In principle many different types of data can be used, there is no restriction that the variable observed can be used directly to deduce model variables, only that the forward process K_n from model variables to observed variables can be done and linearized. Thus for example it would be possible to include satellite radiance data directly in the analysis; work is in progress on a scheme which partially integrates the inversion of such data with the assimilation (Lorenc et al 1985).

c. Balance

The non-linear 'balance' consistent with the observations, which was discussed in 3e, might be improved if the forecast model is specifically altered to selectively damp unrealistic modes. At present we do this by the addition of a term

$$\frac{\partial \underline{y}}{\partial t} = \dots + K \text{ grad div } \underline{y}$$

which selectively damps horizontally divergent motion. More sophisticated notions of balance might be used.

d. Modified successive correction method

The iterative method described in section 2b does not converge to the ideal solution, because the background field forcing $\bar{x}_b - \bar{x}[u]$ is ignored, and because of the approximation in (11). The latter, by having a normalization factor which varies over the analysis grid, changes the structure imposed on the weights by the correlation function. This is alleviated by ad hoc measures to ensure that the normalization factor is smooth. The problems can be avoided by using a modified successive correction method (Bratseth 1985, Lorenc 1986), which calculates a normalization factor in observation rather than grid-point space. Research is in progress into ways of achieving this while still avoiding the need for searching.

e. Quality Control

In theory the iterative analysis method is ideally suited to allowing for the non-Gaussian error distributions which are usually dealt with by quality control algorithms (Purser 1984). However in practice, for operational efficiency, data selection and processing decisions are made for an analysis on criteria which are not ideal for quality control, since quality control requires otherwise redundant data (Lorenc 1984b). Thus there is no intention to include quality control algorithms in the assimilation routine, a prior quality control of all observations will be necessary.

REFERENCES

- Balgovind, R; Dalcher, A; 1983 A stochastic-dynamic model for the
Ghil, M; and Kalnay, E spectral structure of forecast error
statistics.
Mon. Wea. Rev. 111, 701-721.
- Barnes, S L 1973 Mesoscale objective analysis using
weighted time-series observations.
NOAA Tech. Memorandum ERL NSSL-62.
- Bergthorsson, P and Doos, B R 1955 'Numerical weather map analysis'.
Tellus 7, 329-340.
- Bratseth, A M 1985 Statistical interpolation by means of
successive corrections.
Tellus 38A, 439-447.
- Cressman, G P 1959 'An operational objective analysis
scheme'.
Mon. Wea. Rev. 87, 367-374.
- Lorenc, A C 1981 A global three-dimensional
multivariate statistical analysis
scheme.
Mon. Wea. Rev. 109, 701-721.

<p>Lorenc, A C</p>	<p>1984a</p>	<p>Data assimilation by repeated insertion into a forecast model - Principles, practice, problems and plans ECMWF Seminar on data assimilation systems and OSE with particular emphasis on FGGE. September 1984. Met O 11 Tech. Note No. 193.</p>
<p>Lorenc, A C</p>	<p>1984b</p>	<p>Analysis methods for the quality control of observations. Met O 11 Tech. Note No. 194 34pp. ECMWF Workshop proceedings 'The use and quality control of meteorological observations for numerical weather prediction'.</p>
<p>Lorenc, A C</p>	<p>1986</p>	<p>Analysis methods for numerical weather prediction. Quart. J. R. Met. Soc., 112, 1177-1194.</p>
<p>Lorenc, A C; Adams, W; and Eyre, J</p>	<p>1985</p>	<p>The analysis of high-resolution satellite data in the Met Office. Proceedings of ECMWF workshop on high-resolution analysis. Met O 11 Tech. Note 212.</p>

Purser, R J	1984	A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. Preprints, 10th conference on weather forecasting and analysis. Am. Met. Soc. 102-105.
-------------	------	--

1984 A new approach to the optimal
assimilation of meteorological data by
iterative Bayesian analysis.
Preprints, 10th conference on weather
forecasting and analysis.
Am. Met. Soc. 102-105.

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/ICSU 98-111.

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/ICSU 98-111.

Continuous data assimilation at GFDL during FGGE.

Proceedings of ECMWF/WMO Seminar on data assimilation systems and observing system experiments with particular emphasis on FGGE.

3-7 Sept 1984. Reading. Volume 2 pp 125-156.

Continuous data assimilation at GFDL during FGGE.

Proceedings of ECMWF/WMO Seminar on data assimilation systems and observing system experiments with particular emphasis on FGGE.

3-7 Sept 1984. Reading. Volume 2 pp 125-156.

Thiebaut, H J

1985 On approximations to geopotential and
wind-field correlation structures.
Tellus 37A, 126-131.

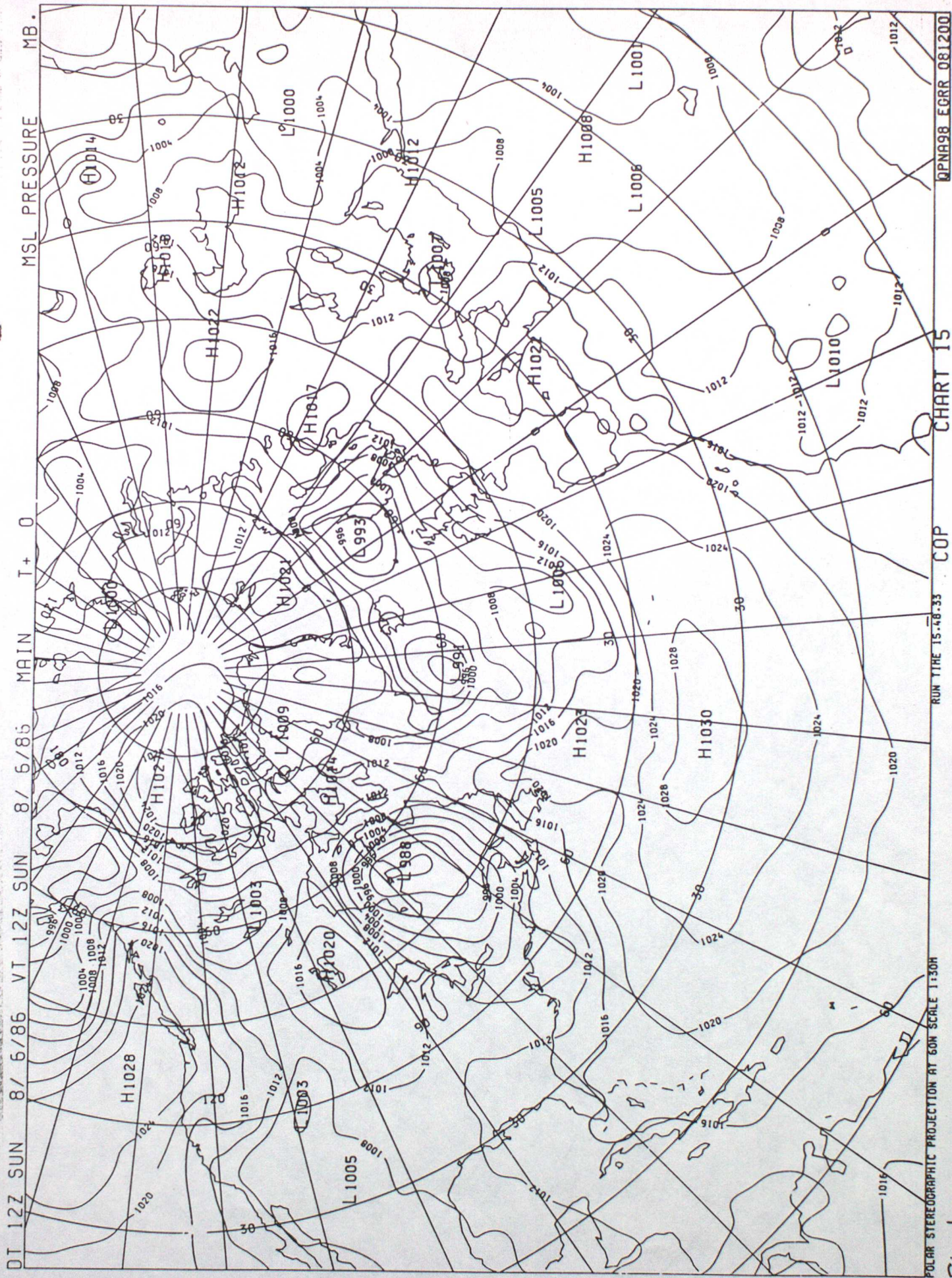


Fig 1a

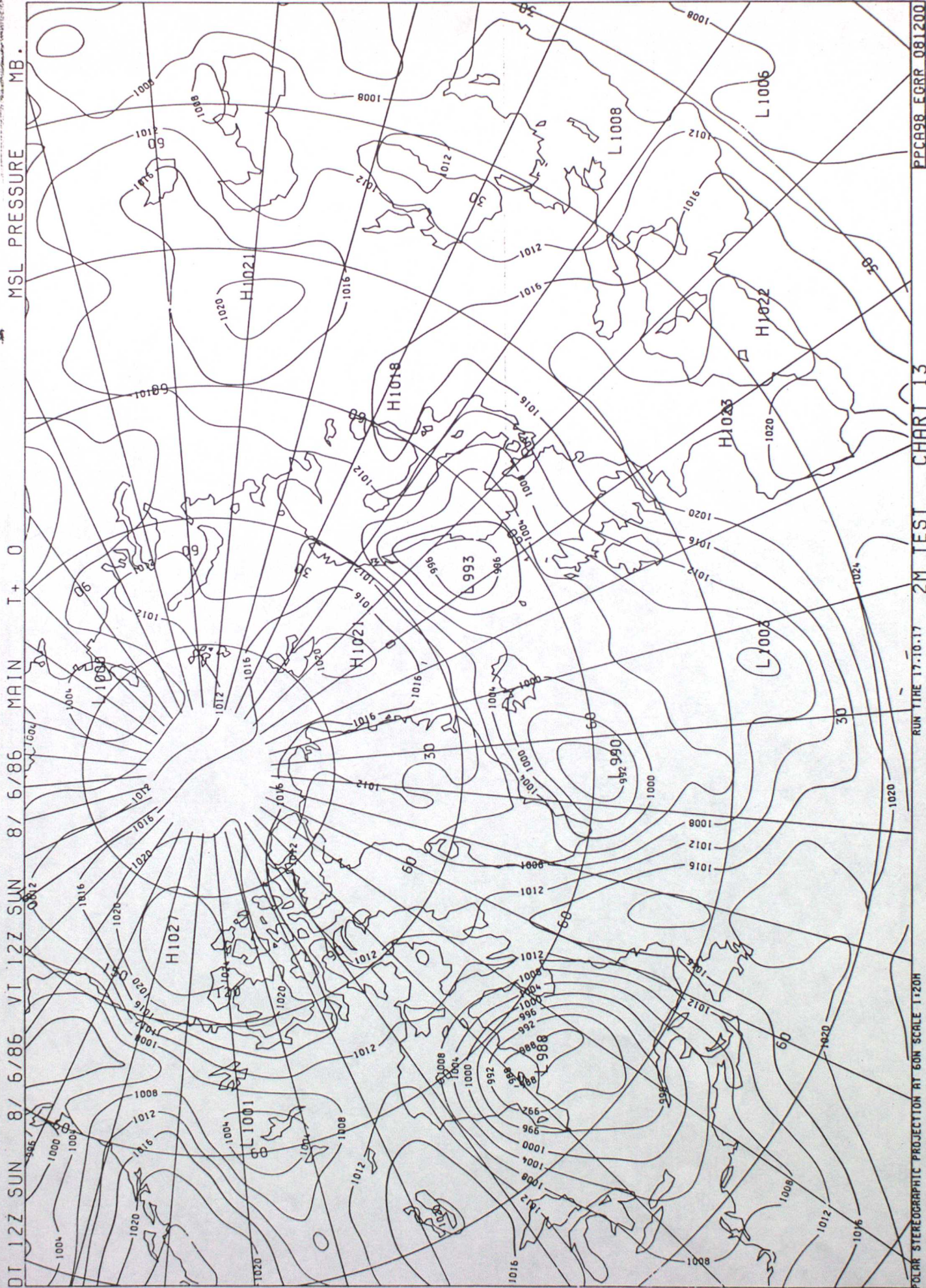


Fig 1b

POLAR STEREOGRAPHIC PROJECTION AT 60N SCALE 1:2000
RUN TIME 17:10:17 2M TEST CHART 13
PCA98 EGRR 081200

500-1000MB
500MB

- THICKNESS DM.
HEIGHT DM.

T+ 0

MAIN

8/ 6/86

SUN

VT 12Z

SUN

8/ 6/86

DT 12Z

SUN



POLAR STEREOGRAPHIC PROJECTION AT 60N SCALE 1:7.7M

RUN TIME 15:48:04

AUXX EGRR CHART 13

PHC450 EGRR 081200

Fig 2a

500-1000MB
500MB



RUN TIME 17.11.01 3M TEST

CHART 15

Fig 2b



Fig 3a

MAX WIND SPEED KT.

T + 0

MAIN

8/ 6/86 VT 12Z SUN 8/ 6/86

01 12Z SUN



POLAR STEREOGRAPHIC PROJECTION AT 60N SCALE 1:250,000

RUN TIME 17:12:43

2M TEST CHART 37

QWNA96 EGLL 081200

Fig 3.b



DT 12Z SUN 8/ 6/86 VT 12Z SUN 8/ 6/86 MAIN T+ 0 - - THICKNESS DM. 500-1000MB
 HEIGHT DM. 500MB



POLAR STEREOGRAPHIC PROJECTION AT 60N SCALE 1:30M
 RUN TIME 16.23.27 HQSTC CHART425 PHMSO EGRR 081200

Fig 5a



DT 12Z SUN 8/ 6/86 VI 12Z SUN 8/ 6/86 MAIN T+ 0

500-1000MB
500MB

THICKNESS DM.
HEIGHT DM.

PHASO EGR 081200

CHART 425

POLAR STEREOGRAPHIC PROJECTION AT 60N SCALE 1:30H

Fig 5b



CM TEST CHOPT/CE

Fig. 6b