TECHNICAL MEMORANDUM

15

Modifications to the operational analysis suite

P.Lönnberg

Research Department

1980

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

A large set of modifications and improvements to the data assimilation system has been prepared during the last months. A number of parallel runs with slightly different versions have been carried out. Section 1 describes the modifications implemented operationally in the middle of March. Section 2 contains the tasks to be completed in the near future with some preliminary results. Large savings in computer time are then expected.

1. MODIFICATIONS IMPLEMENTED IN MID-MARCH

This Section describes the modifications that were incorporated operationally from 11 March 1980 to the analysis system. The tests carried out show a marked improvement in the analysis of small scale phenomena (see Figs. 1 and 2 ; the proposed version is the new operational system and the "current" operational was the version used before the changes). The fit to the observations is also slightly better than before. As a result of allowing and forcing in more observations, a minor increase in the computer time has been noticed. However, the removal of artificial limits facilitates future refinements of the data selection. The changes can be grouped into six parts:

a) data checking, selection and modification

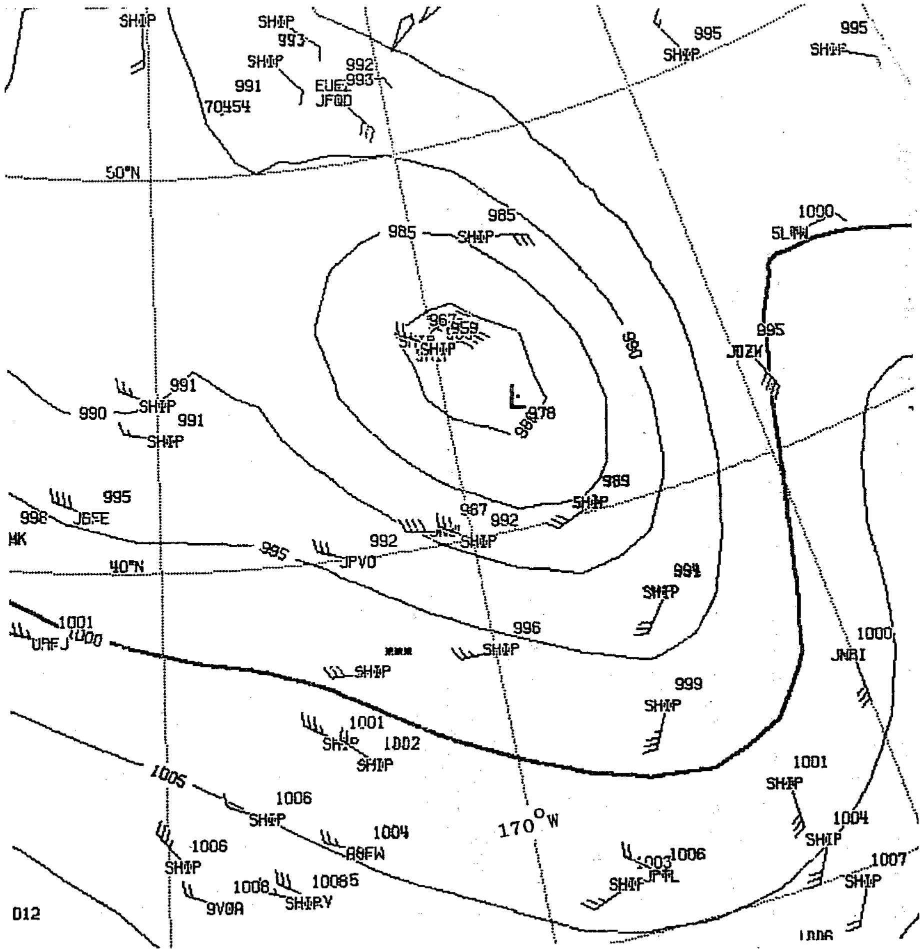
b) super-observation formation

c) use of off-time observations

d) statistics

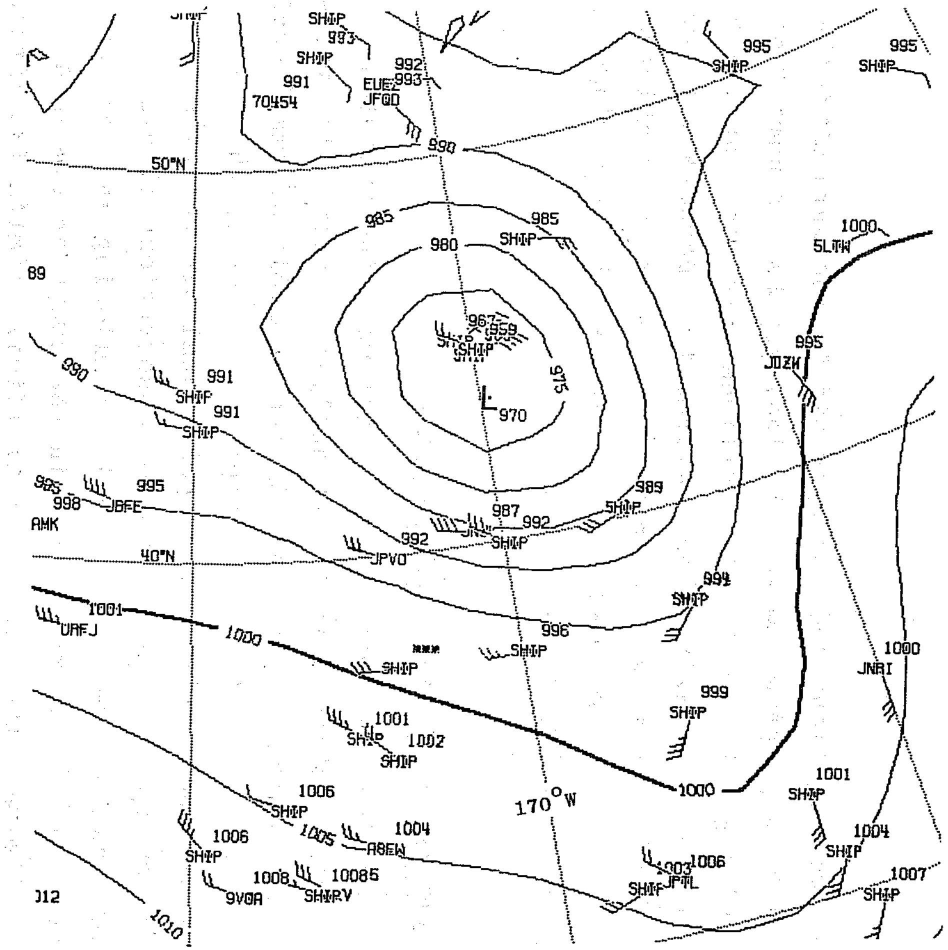
e) stratospheric analysis

f) new version of the forecast model and removal of the filtering of initialised fields (operational since 29 February 1980)



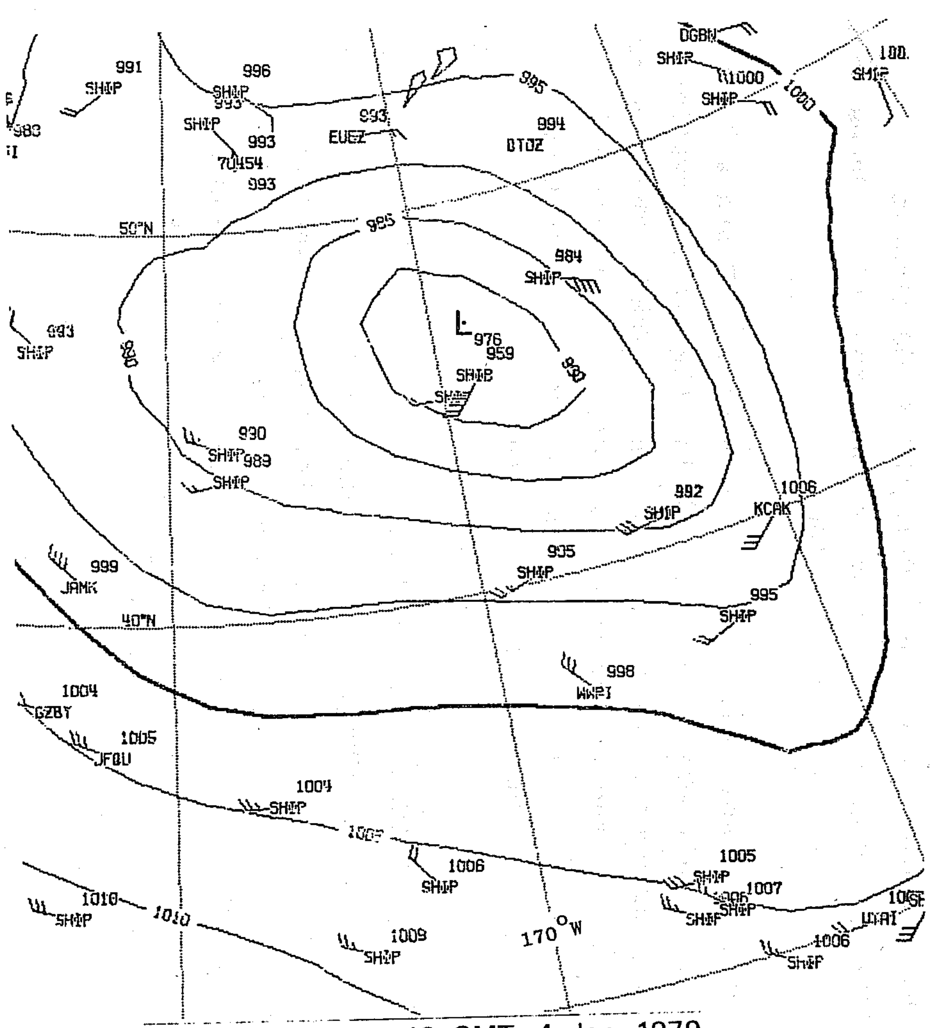
Sea level pressure; 06 GMT; 4 Jan. 1979

Fig. 1a Current operational version (Experiment P1)



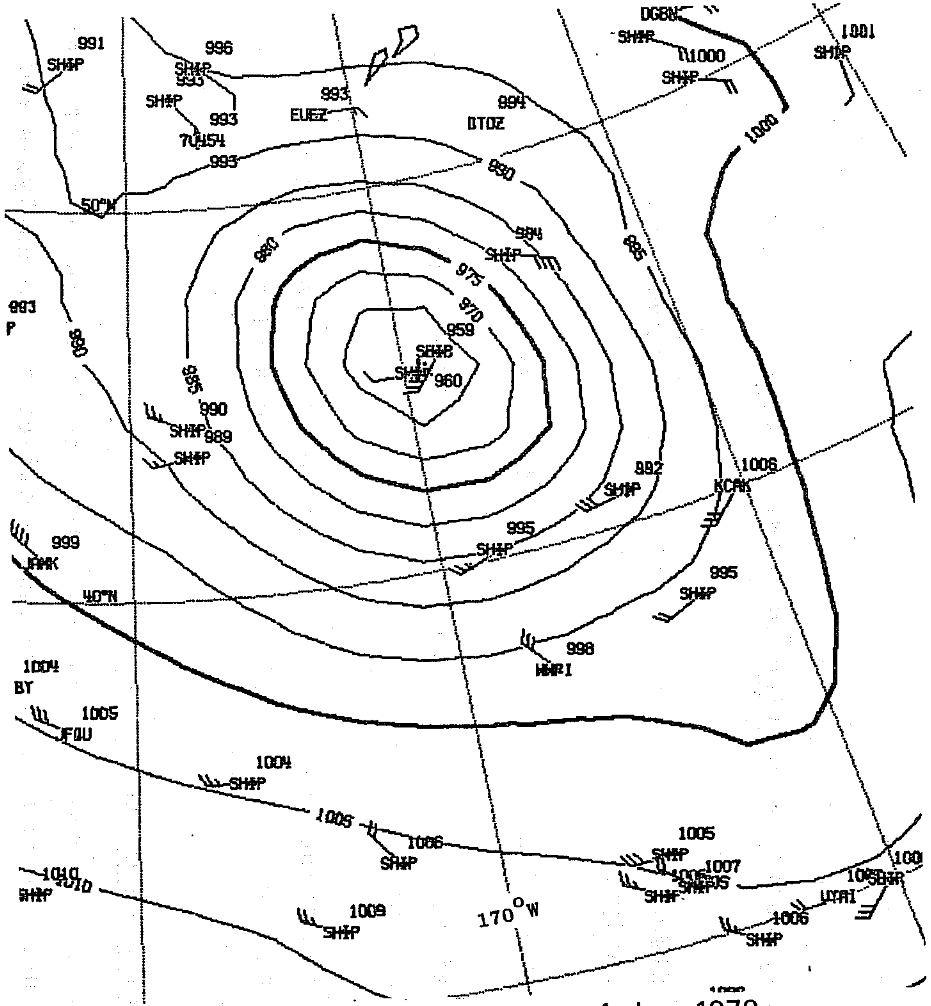
Sea level pressure; 06 GMT; 4 Jan. 1979

Fig. 1b Proposed version (Base-line Experiment B4)



Sea level pressure; 12 GMT; 4 Jan. 1979

Fig. 2a Current operational version (Experiment P1)



Sea level pressure; 12 GMT; 4 Jan. 1979

Fig. 2b Proposed version (Base-line Experiment B4)

1.1 Data checking, selection and modification

These changes aim at

a) discarding observations that would degrade the analysis,

b) improving the quality of the data used,

c) giving a larger flexibility in rapidly changing situations,

d) improving the final selection of data, especially as regards vertical continuity and analysis matrix size.

1.1.1 Discarding of observations

a) Observations from the same station and location but different times have usually had an unfavourable effect on the analysis due to super-obbing. This has been the case for SYNOP observations which can be reported every hour. Only the SYNOP closest to analysis time is retained. In an operational run this reduces the number of observations by more than 1000.

b) A sounding is sometimes reported both as a TEMP and a PILOT. In order to avoid the effect of super-obbing the two observations are merged giving precedence to the data in the TEMP. The merging takes place only if the observation times differ by less than 30 minutes. In a typical FGGE assimilation (00 or 12Z) about 200 pairs of co-located TEMP/PILOTs could be identified.

1.1.2 Improvements to the data quality

a) For an off-time SYNOP/SHIP/SHRED the mean sea level pressure is corrected by the reported tendency. For a moving ship the tendency is corrected by the displacement againt the pressure gradient. The gradient is computed from the observed wind using the assumptions that the crossisobar angle is 20° and that the surface wind speed is 80% of the geostrophic wind.

The error in the observation due to asynopticity (see Section 1.3) is reduced by a factor of 4 if all data used for correcting the pressure were flagged as reliable. If the tendency correction is considered less reliable (data flagged or missing wind and/or ship movement) the error growth is reduced by a factor of 2.

b) The mean sea level pressure is checked for 100 mb error in SYNOPS. If P_s differs by more than 30 mb and (P_s ± 100 mb) by less than 30 mb from the first-guess, the surface pressure is changed by 100 mb.

1.1.3 Changes in flagging of data

a) The error limits in the comparison of the observed value against the first-guess have been too harsh especially as regards rapidly developing situations. Instead of setting flags at normalised observation deviations of 3, 4 or 5 times the prediction error, the limits are increased to 4, 6 and 8.

b) The method of checking data by solving the analysis equation to the data point has some errors due to the method itself and the assumed statistics. The estimated interpolation error variance has thus been increased by 10% of the estimated forecast error variance in the data checking.

c) In the present operational version close data which do not agree in the "buddy" check have their flags increased by one. This can exclude the data from being used at a later stage. Only a clear disagreement should lead to an increased flag. A separate test to detect discrepancies in the data is proposed. This includes the variations due to separation in time and space in the first-guess error correlation:

$$\mu^p_{ij} = < \alpha^p_i \ \alpha^p_j > = 1 - \tfrac{1}{2} r_{ij}^2/b^2 - \tfrac{1}{2} t_{ij}^2/c^2$$

α^p is the normalised prediction deviation from the true value, b is the horizontal scale (500 km) and c the time scale (4 hours). The test for disagreement is then

$$(\delta^o_i - \delta^o_j)^2 \geqslant \mathrm{BUDDIS}\ (\varepsilon^{o2}_i + \varepsilon^{o2}_j + r_{ij}^2/b^2 + t_{ij}^2/c^2)$$

BUDDIS is set to 16. δ^o and ε^o are the normalised observation deviation and error, respectively.

1.1.4 Data selection in main analysis program

a) An observation in the central box or its near neighbours is always selected provided that the maximum number of observations has not been reached. This forces in data which otherwise would not have been used due to sufficiency of data and lead to a more even distribution of data.

b) The data sufficiency test (10 data) in the final stage has been applied separately for each level and variable independent of the amount of data at nearby levels. The test has been changed to stop the selection of data for a specific level and variable when the sum of data from that level and the level below and above is 3*10. This allows for a varying vertical data density.

c) To avoid vertical discontinuities the data volume is, if possible, extended by one level in both directions. This means that data is always included from levels outside the analysis slab.

d) Useful data from outside the analysis slab form a supplementary set to the inner data. The amount of supplementary data for a certain level and variable has been linearly dependent on the vertical correlation between that level and the closest level in the slab. Surface data could therefore be used in the stratospheric analysis. The dependency is changed to the square of the vertical correlation. This has a substantial effect on the matrix size and consequently on the computer time.

e) When selecting observation for an analysis box, the buffer limit of 50 observations has often been reached. The value is increased to 100.

A small increase of the matrix size results from this tuning.

1.2 Super-observation formation

As the number of observations strongly affects the computer time and the behaviour of the analysed fields near the box boundaries some improvements to the super-obbing have been coded:

a) The prediction error correlation has been assumed to be 1 when forming a "super-observation". This assumption can be justified when the distance between the observations is small or the "super-ob" is formed between the observations. When the "super-ob" is formed at one of the observation points the differences in space and time are included:

$$\delta_s^o = \frac{\delta_a^o \, (\varepsilon_a^{o2} + 1 - \mu_{ab}^{p2}) + \delta_b^o \, \varepsilon_b^{o2}}{\varepsilon_a^{o2} + \varepsilon_b^{o2} + 1 - \mu_{ab}^{p2}}$$

δ^o and ε^o are the normalised observation deviation and error, respectively. Subscript s indicates the "super-observation". In the above formula the "super-ob" is formed at the position of observation a. The calculation of the correlation μ_{ab}^p was described in Section 1.1.3 c.

b) The "super-observations" have been formed without regard to the observation time. In the case the two observations are considered equally important (e.g. same type), the separation in time is checked. If they are less than one hour apart or their absolute differences to analysis time differ by less than 30 minutes, "super-obbing" takes place. Otherwise the observation closest to analysis time is retained and the other is discarded.

Because of the improvements from a and b the "super-obbing" radius is increased from 100 km to 150 km. The "secondary" observation formation radius has accordingly been set to 190 km (previously 150 km). This reduces slightly the analysis time and the discontinuities at the box boundaries.

1.3 The effect of asynopticity in the estimate of the error of an observed value

The error assigned to a datum consists of two effects; the error of the observing method and an estimate of the persistence error from the difference between analysis and observation time:

$$E^{o} = \sqrt{E_{m}^{2} + E_{p}^{2}}$$

The maximum 24 hour persistence error growth is as follows:

	1000-700 mb	699-250 mb	249- 0 mb
U,V	9 $^{m}/_{s}$	18 $^{m}/_{s}$	27 $^{m}/_{s}$
Z	48 m	60 m	72 m
T	6 K	7 K	8 K

The season and latitude are accounted for by

$$a = \sin\left(2\pi \frac{\text{day}}{365.25} + \frac{\pi}{2}\right)$$

$$b = 1.5 + 0.5\ \text{MIN}\ [\text{MAX(LAT}, -20), 20]\ /20*a$$

The growth rate is then computed from the maximum growth rate for a time difference Δt:

$$E_{p} = \frac{E_{pmax}}{6}\ (1 + 2\text{SIN}|\phi|)*b*\Delta t$$

1.4 Changes in assumed statistics

a) New covariances

The previously used vertical correlation matrix of the first-guess contained some unrealistically large values for the height error between low tropospheric levels and 10 mb. New correlations have been modelled assuming a decrease as exp $\left(\frac{Z_1 - Z_2}{H}\right)$, where H is scale height of the errors. H is 3 km at 1000 mb and 8 km at 10 mb in mid-latitudes. In the tropics H is 2 and 8 km, respectively (see working paper by A. Hollingsworth).

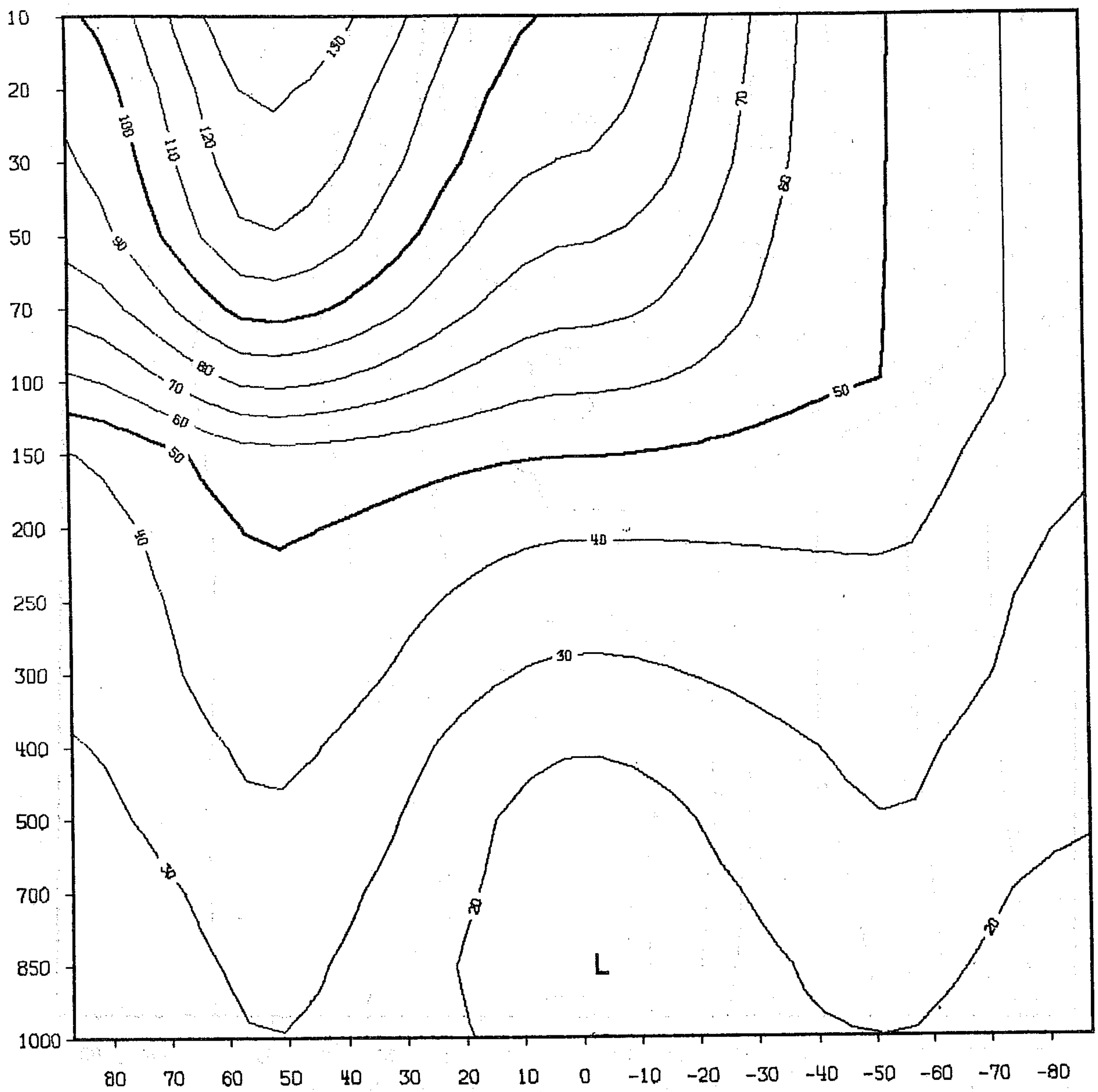


Fig. 3 Mean six hours height forecast error (m) for January

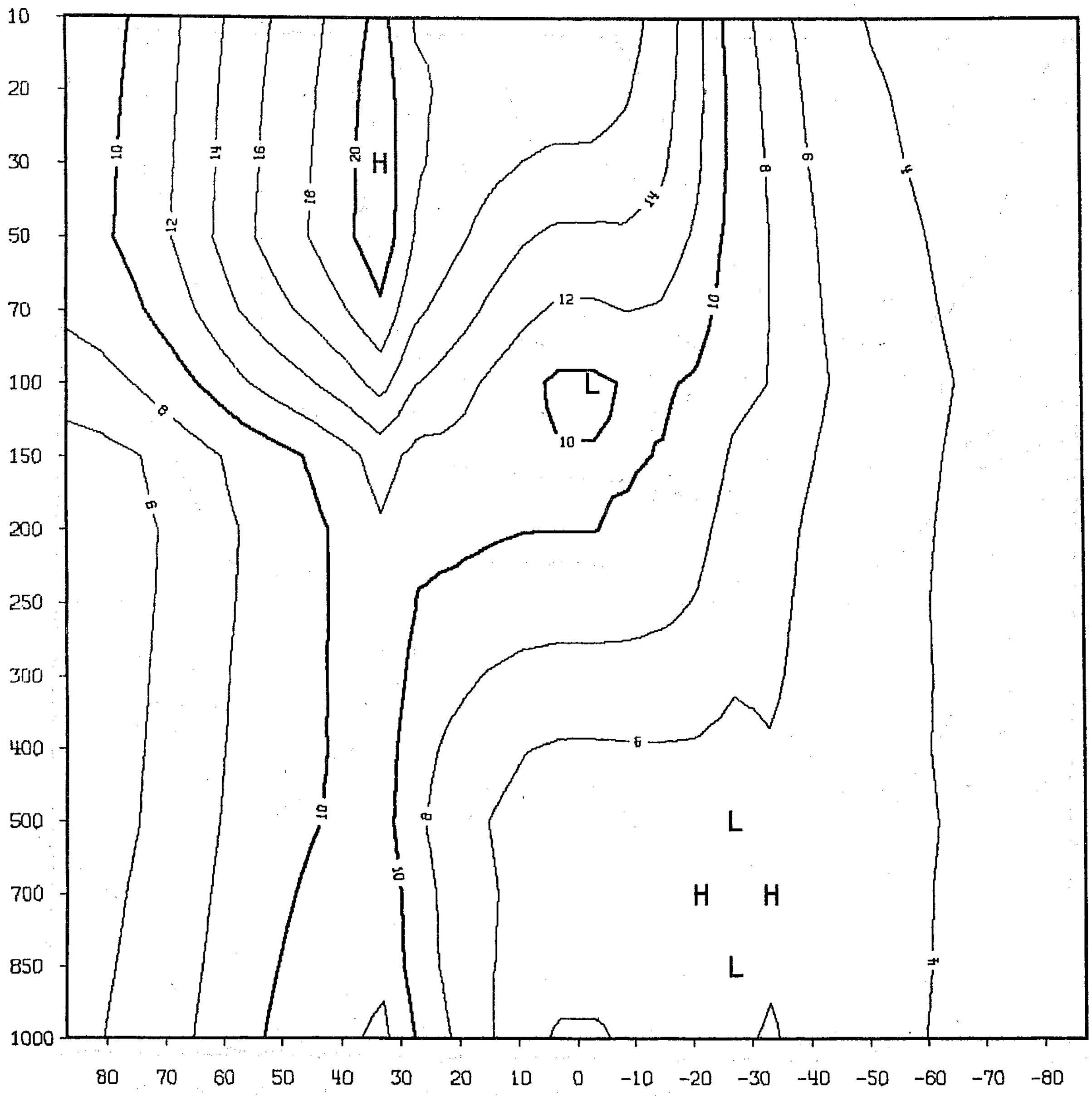


Fig. 4 Mean six hours u,v forecast errors (m/s) for January

A new set of correlations for the wind errors in the tropics has also been compiled.

The errors of the forecasted heights have been increased significantly. The January values are shown in Fig. 3 and the corresponding errors for the wind components in Fig. 4. A new set of covariances has been derived using the modified correlations and errors and a relaxed streamfunction-height correlation (from ±0.95 to ±0.85 polewards of 30^o latitude).

The impact of the new covariances can be seen in Fig. 5.

b) Larger vertical variation of horizontal width of structure function

The structure function used to calculate the correlation between the heights at points with horizontal coordinates $\vec{r}_i$ and $\vec{r}_j$ and vertical coordinates p_i and p_j is $\exp(-\frac{1}{2}(\vec{r}_i - \vec{r}_j)^2/b^2)$ where b is the width of the structure function. It is calculated as an average of $b(p_i)$ and $b(p_j)$, which are obtained by prescribing them at 3 levels (1000, 200 and 10 mbar) and by linear interpolation in between. A larger vertical variation of b than previously has been assumed:

	old values	new values
b(1000 mbar)	500 km	250 km
b(200 mbar)	600 km	500 km
b(10 mbar)	700 km	750 km

The decrease of b at the surface gives a finer scale to the analysed fields as illustrated in Figs. 6 and 7. The experiments in Figs. 6 and 7 also have different data selection schemes and forecast error variances.

The stronger vertical variation of b can cause the computed correlation matrix to be ill-conditioned. A scheme to detect (and cure in a meteorologically acceptable way) such ill-conditioning has been developed (see working paper by G. Cats and D. Robertson).

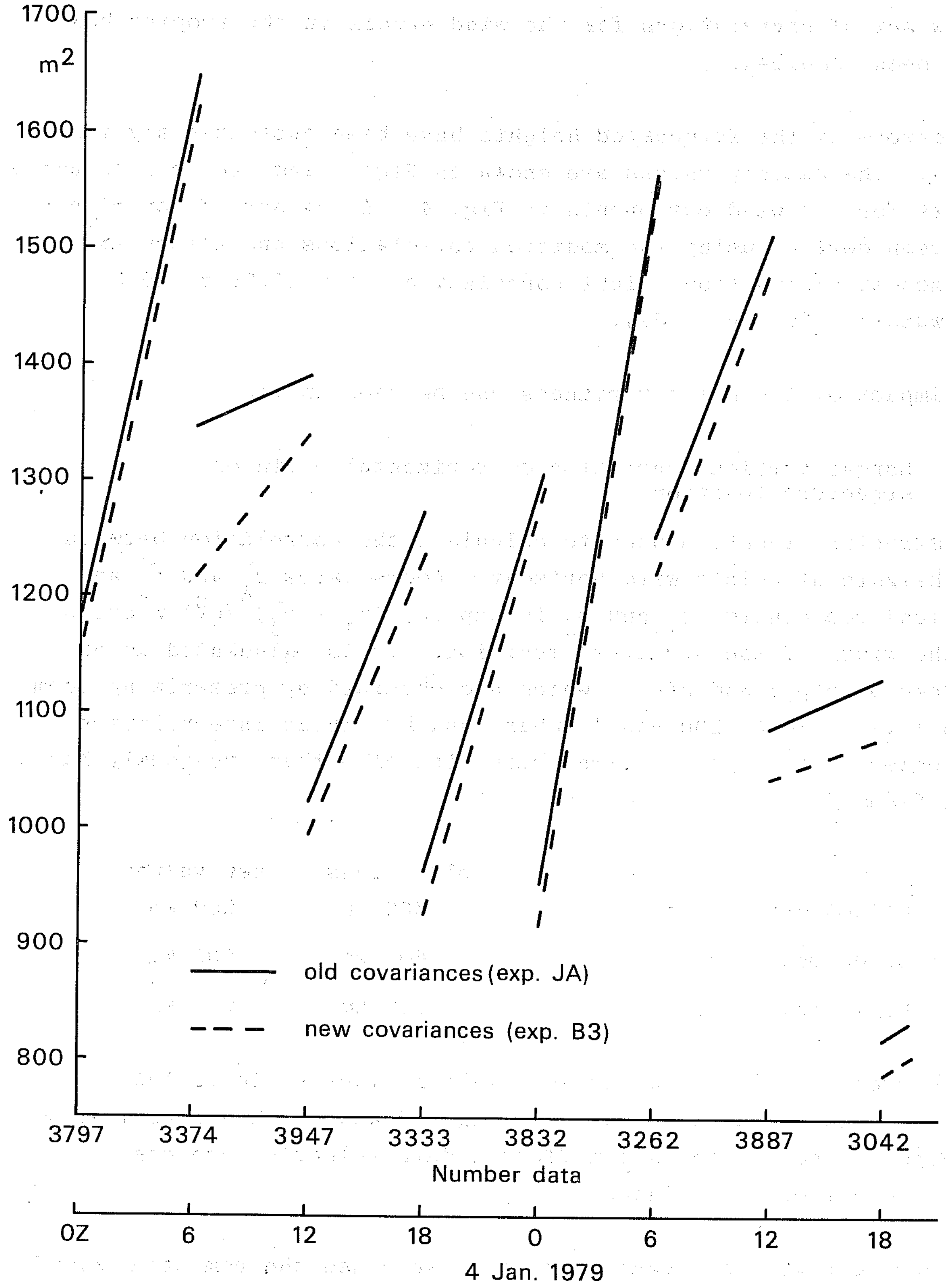


Fig. 5 Mean square deviation (m^2) of analyses and 6 hour forecasts for all 1000 mb height and PMSL data

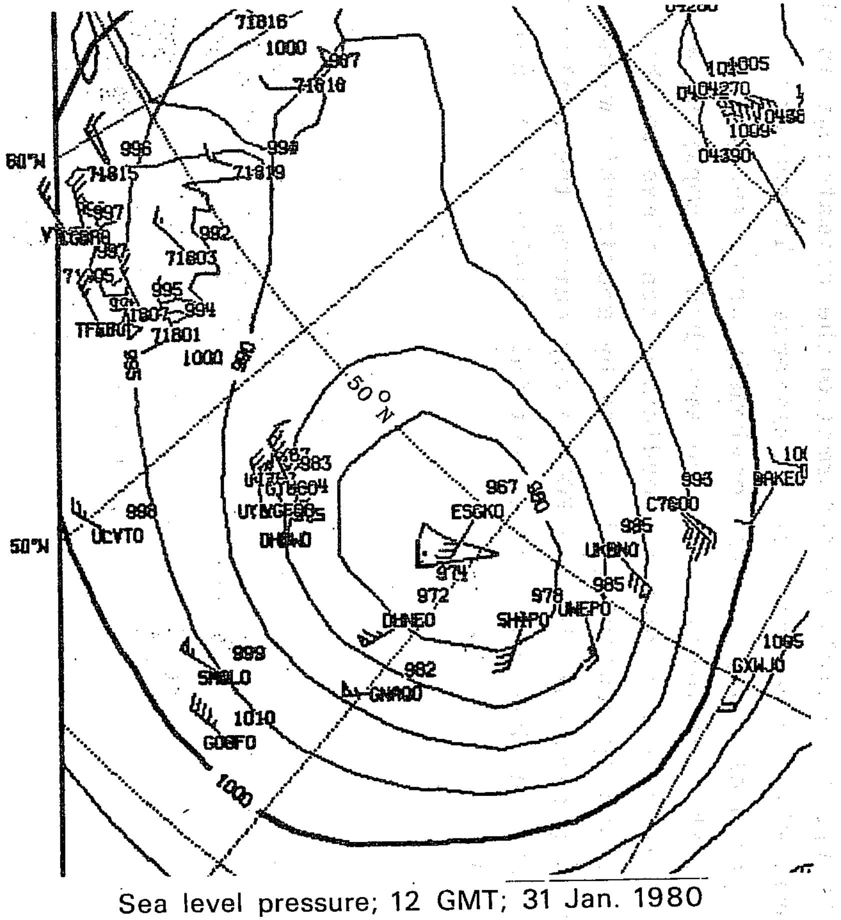


Fig. 6 OLD values of b

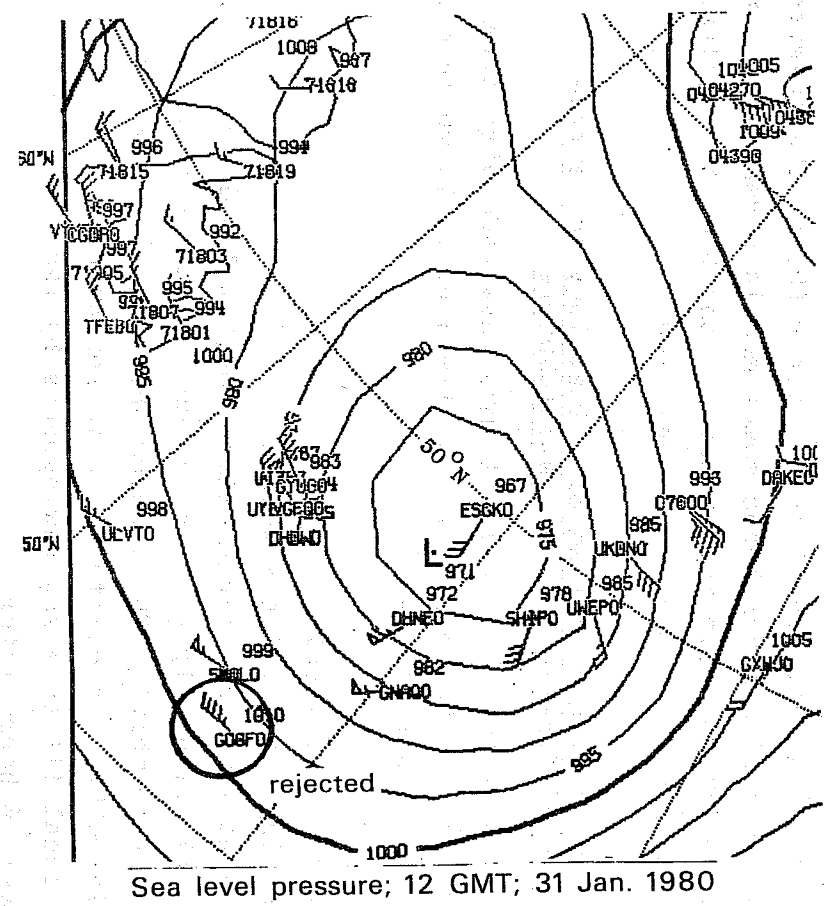


Fig. 7 NEW values of b; different data checking

c) The use of the horizontal scale factor (b) of the observation box and not the analysis box leads sometimes in a region of rapid change in b to ill-conditioning. Instead, the b of the analysis box will be applied to all observations used.

1.5 Use of persistence instead of climatology for extrapolation of first-guess into stratosphere

Figs. 8 and 9 show two consecutive 10 mbar analyses with climatology used to extrapolate the model forecast field up to 10 mb in order to get the first-guess field:

> first-guess at level p =
> climatology at level p + weight *
> (model at level 50 mbar - climatology at level 50 mbar), with weight = $\frac{3}{4}$ at 30 mbar, $\frac{1}{2}$ at 20 mbar and $\frac{1}{4}$ at 10 mbar). The figures show that the low initiated by a satellite track is replaced by a very local fit to a few rawinsondes by the analysis 6 hours later.

To carry through the satellite information the climatology in the above formula will be replaced by persistence. To prevent persistence from carrying through obviously wrong analyses indefinitely, the persistence will be smoothed by applying a 1-2-1 filter in both horizontal coordinates ten times to the height field, and by replacing the winds by geostrophic winds. This scheme has been used by the FGGE group during the analyses of January 1979 and has produced satisfying results up to now (see Fig. 10).

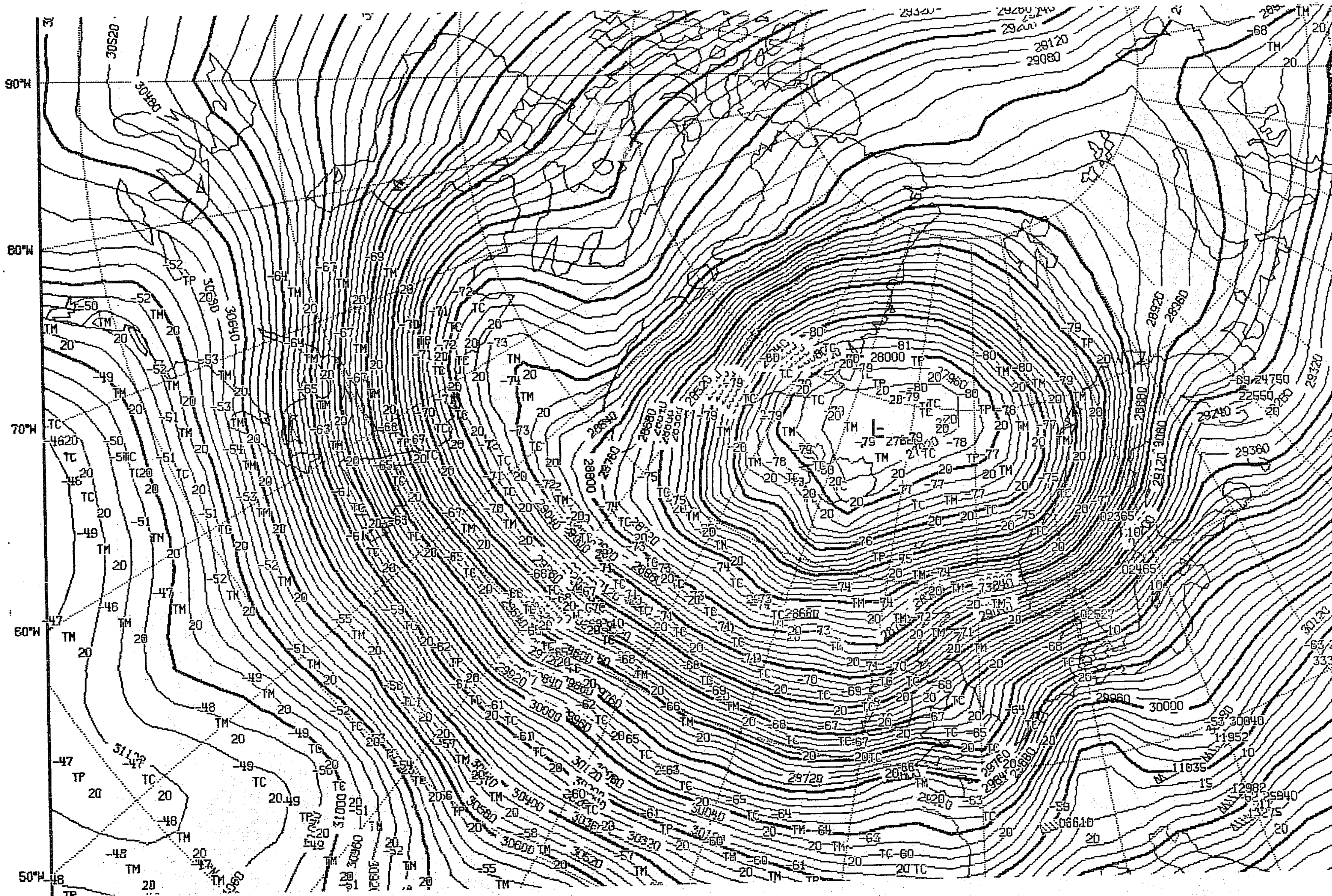


Fig. 8 Analysis; 10 mb; 1 Jan. 1979; 12 GMT

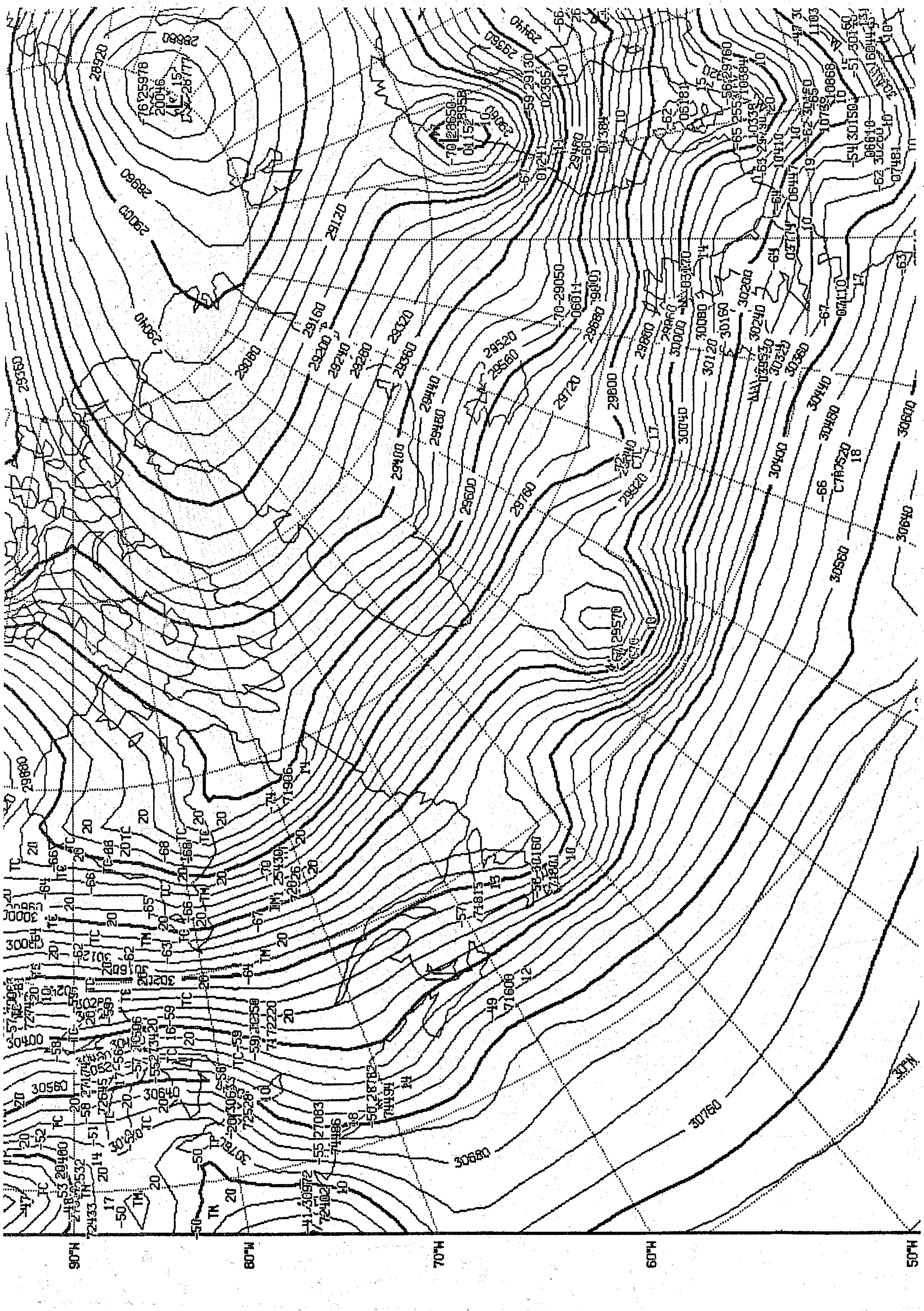


Fig. 9 Analysis; 10 mb; 1 Jan. 1979; 12 GMT. Climatology used to get first guess

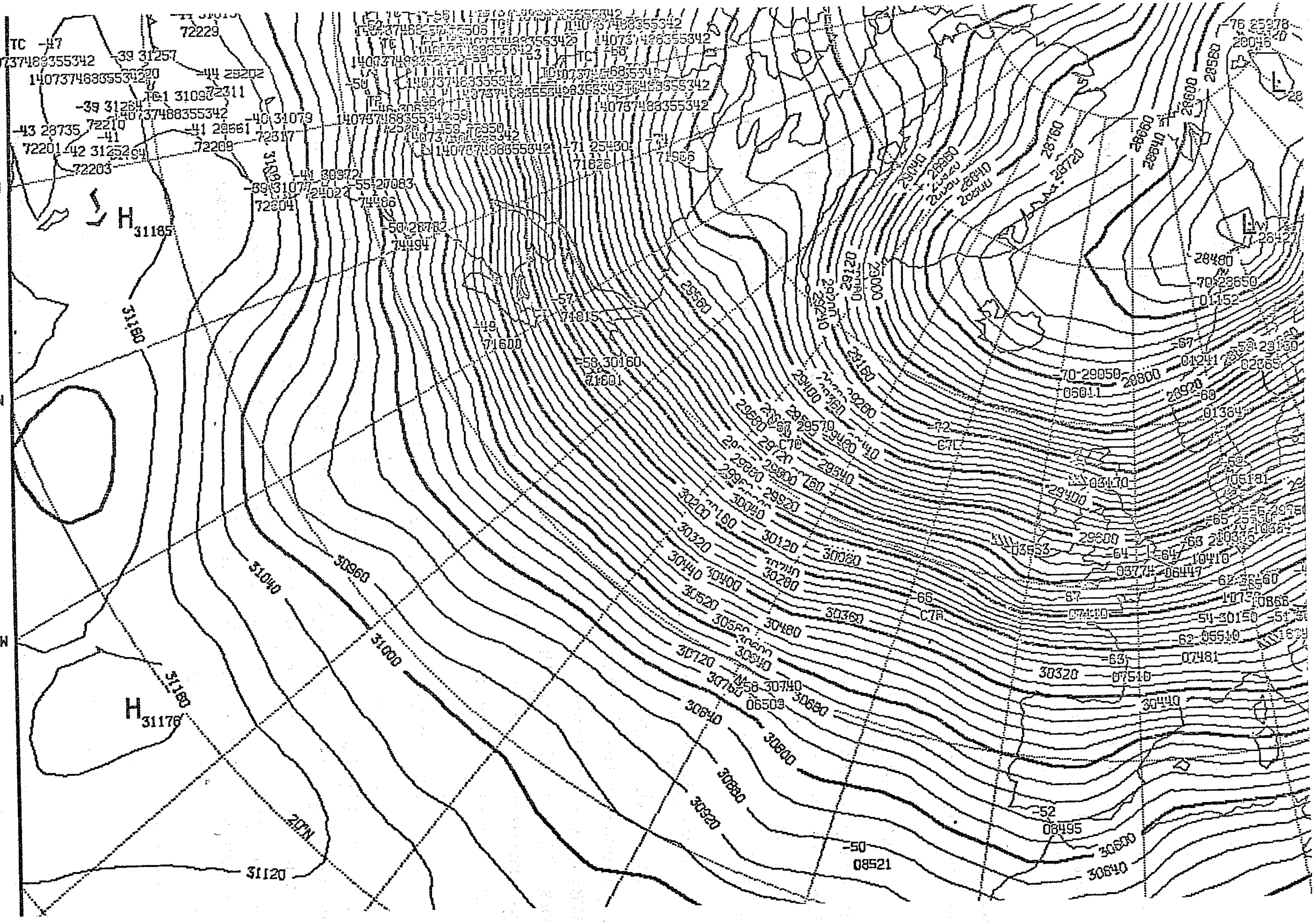


Fig. 10 Analysis; 10 mb; 1 Jan. 1979; 12 GMT

2. METEOROLOGICAL CHANGES EXPECTED IN THE NEAR FUTURE

2.1 Interpolation of increments

A considerable reduction in the interpolation error has been obtained by interpolating ($\sigma \leftrightarrow p$) the analysed increments instead of the full fields. The loss of mass (up to 2 mb/cycle) which occurs at present can be avoided.

The structure of PBL will be better preserved. The modifications will be implemented in the near future.

2.2 Data selection

The analysis matrix contains a large amount of data with marginal effect on the analysed increments. Some tests will be made to reduce the number of data without significantly degrading the results. The expected saving of computer time could be substantial.

2.3 Overlapping boxes of multiple gridpoint analysis

A significant reduction of the box boundary jumps can be achieved by extending the analysis volume into neighbouring boxes and merging the analysed gridpoint values in the overlap areas. Figs. 11 and 12 show that the jumps disappear while the gradients are preserved. For each analysis of the same gridpoint an associated weight is calculated:

$$W_k = (1 - 2\ |x - x_k|/(cL_x))*(1 - 2\ |y - y_k|/(cL_y))$$

(x,y) and (x_k,y_k) are the coordinates of the gridpoint and the central point of the analysis volume k, respectively. L_x and L_y are the sidelengths of the box. c defines the extension of the analysis volume into its neighbours; $c > 1$. The normalised increments (α) for the gridpoint (x,y) is obtained by taking the weighted average of the increments (α_k) of the influencing analysis volumes:

$$\alpha = \Sigma\alpha_k\ W_k/\Sigma W_k$$

By some tuning at the overlap in conjunction with a less critical data selection, it is hoped that the elapsed time can even be slightly reduced.

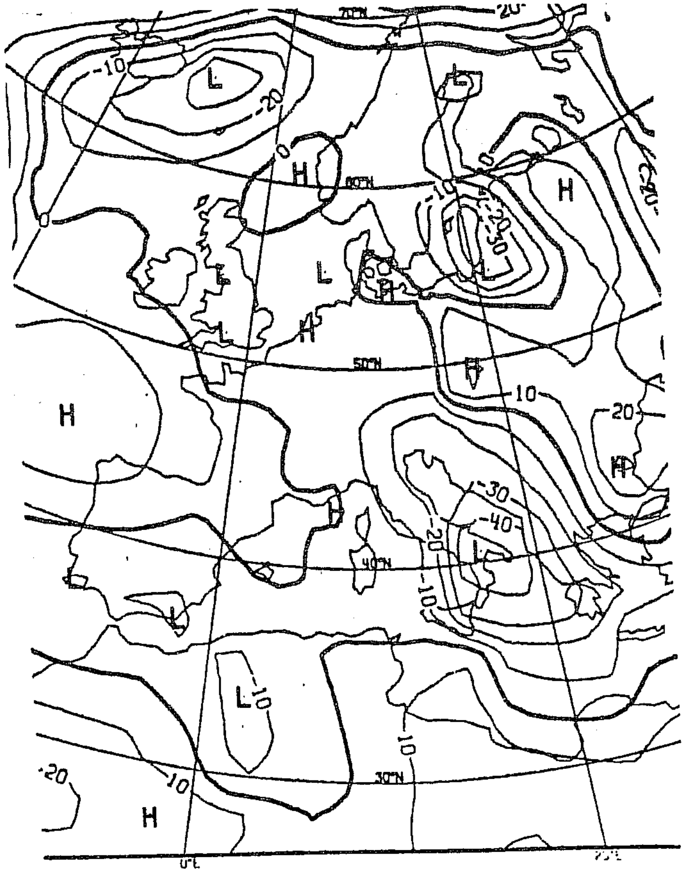


Fig. 11 Base-line Experiment B4; 500 mb height field; analysis increments

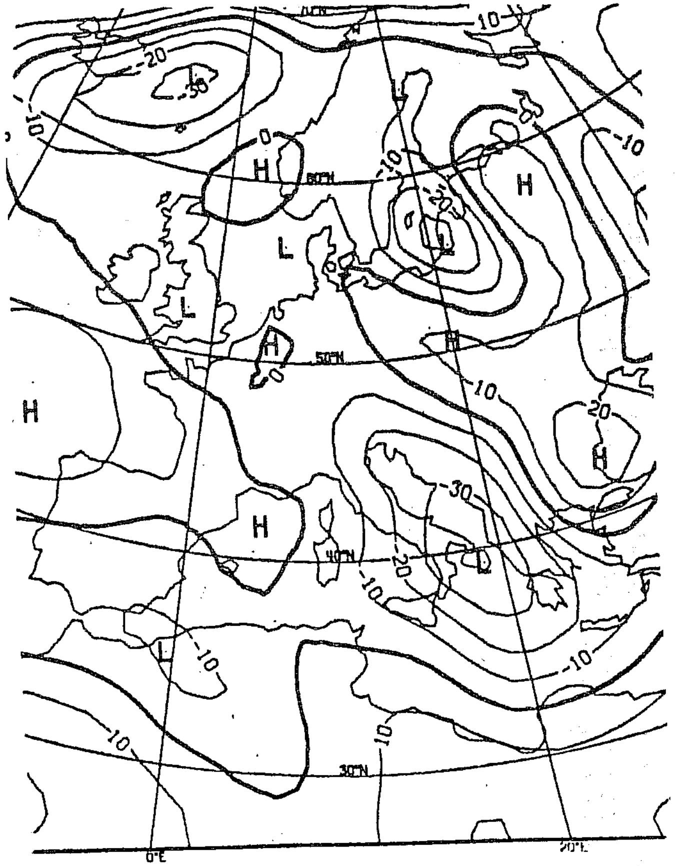


Fig. 12 Experiment J2; overlapping boxes (B=2) + quick data checking

References

Cats, G. and D. Robertson, 1980: Condition of the covariance matrix and stability of its inversion algorithm. ECMWF Working Paper No. 1/6/E/GJC/071/1980.

Hollingsworth, A., 1979: Review of operational prediction errors for height. ECMWF Working Paper No. 1/26/E/DD/069/1979.