TECHNICAL MEMORANDUM

94

Evaluation of analysis increments at model levels

P. Undén

Research Department

December 1984

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

Abstract

In order to obviate the interpolation of analysis increments from standard pressure levels to model hybrid levels, a method has been developed for calculating these increments directly at the desired model levels. It utilizes continuous vertical structure functions so that the optimum interpolation formalism can be applied directly. The method works well and improves the 6 hour first-guess forecasts in the data assimilation but has a rather neutral impact on the 10-day forecasts.

The method described in this report was implemented into operations on 13 November 1984.

C O N T E N T S

1. INTRODUCTION

The previous ECMWF data assimilation system used data at, or interpolated to, standard pressure levels and the analysis increments were evaluated at these pressure levels. The pressure level increments were then interpolated to the model hybrid levels. (For details see the ECMWF Data Assimilation Scientific Documentation). It has been suggested that it would be preferable to use the optimum interpolation (O.I.) scheme to evaluate the increments directly at any desired model levels (Hollingsworth, 1982). Such increments would then be consistent with the constraints built into the multivariate O.I. scheme and would also be more accurate, at least when enough data are present to define vertical derivatives of the model variables. However, this approach relies on the assumption that a continuous functional representation can be found for the vertical correlations and forecast error variances. The analysis increment a_k at a gridpoint (or level) can be written as

$$a_k = (\underline{d}^T \cdot \underline{\underline{M}}^{-1})\underline{P}_k \qquad (1)$$

where $\underline{\underline{M}} = \underline{\underline{P}} + \underline{\underline{O}}$ is the O.I. matrix representing forecast error correlations between data points ($\underline{\underline{P}}$) plus observational error correlations ($\underline{\underline{O}}$); $\underline{d}$ is a column matrix containing the data, and when multiplied by $\underline{\underline{M}}^{-1}$ it gives a vector of coefficients $(\underline{\underline{d}}^T \cdot \underline{\underline{M}}^{-1})$. The analyses increments are then given by multiplying this vector by $\underline{P}_k$ which is a column vector of correlations between the data points and the model gridpoints (and levels); it is these correlations that rely on the continuous representation when model level variables are evaluated.

The continuous representation of forecast error variances is needed to calculate correlations with thickness data and for renormalisation of the analysis increments at model levels.

2. CONTINUOUS VERTICAL CORRELATIONS AND FORECAST ERRORS

Various functions have been tried for representing vertical forecast error correlations. One problem is that the structure of the correlations varies considerably between the stratosphere and the troposphere. Furthermore the 'tails' of the correlations (i.e. the far away correlations) are rather uncertain as they depend a lot on the separation of observational errors. It was therefore decided to use a simple exponential function augmented by the transformation suggested by Cats (1982) which only takes error correlations between adjacent levels into account when constructing the full matrix of modelled correlations.

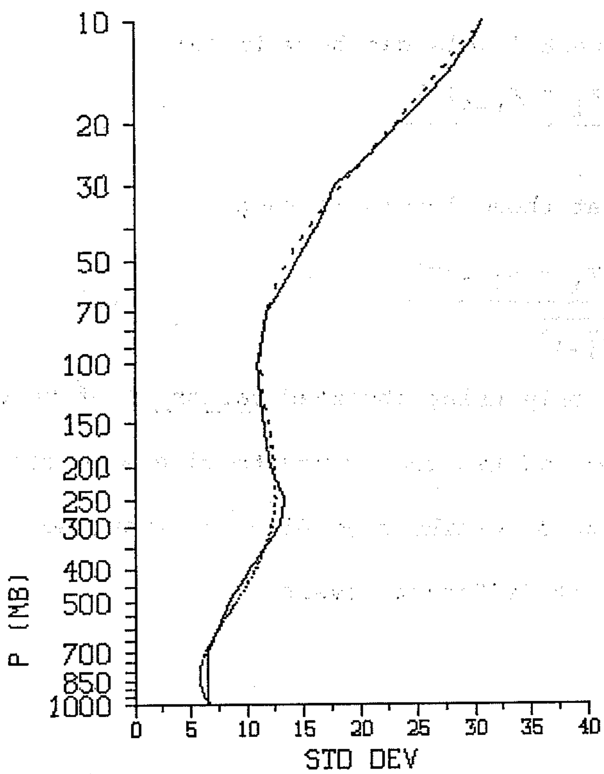
2.1 Forecast errors

The forecast errors of heights and winds needed for normalisations are fitted by using polynomials in ln(p) mainly due to their computational efficiency. However a polynomial of rather high degree is needed to get an adequate fit. When fitting data from 15 pressure levels a sixth degree polynomial was needed to give a reasonable representation (see Fig. 1).

2.2 Vertical correlations using an exponential function with Cats' transformation

Cats' transformation is a way of using the information about observed thickness errors to define the height error correlations between all pairs of adjacent levels.

An exponential correlation model is chosen, fitted to the set of adjacent levels and the model then defines the other correlations. Other sets of levels could also be used but the adjacent ones are likely to give the most stable thickness statistics and are also the most important to have accurately represented for the analysis. The model parameter will then describe a



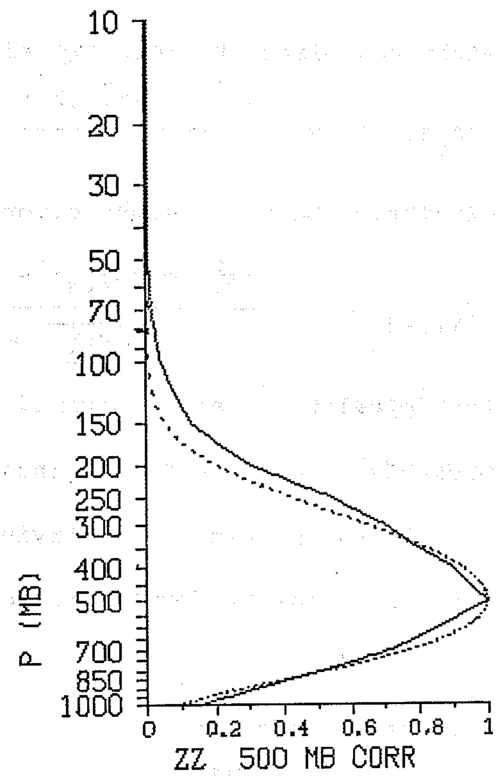


Fig. 1 The height forecast error variance (full line) and the fitted function (dashed).

Fig. 2 The vertical height error correlation between 500 mb and other levels. The dashed line is the fitted function using eq.(4).

unique transformation of the vertical coordinate and the correlations will give symmetric correlation matrices. The transformation has the effect of compressing or expanding the vertical coordinate depending on the correlation scales.

The height covariance between two adjacent levels can be written:

$$\langle z_i z_{i-1} \rangle = \frac{\langle z_i^2 \rangle + \langle z_{i-1}^2 \rangle - \langle (z_i - z_{i-1})^2 \rangle}{2} \qquad (2)$$

The correlation between height errors at those levels is then

$$C_{i,i-1} = \frac{\langle z_i^2 \rangle + \langle z_{i-1}^2 \rangle - \langle (z_i - z_{i-1})^2 \rangle}{2\sqrt{\langle z_i^2 \rangle} \; \sqrt{> z_{i-1}^2 >}} \qquad (3)$$

Thus the correlation can be derived by only using observed <u>variances</u> of height and thickness forecast errors. This method has been found to give more stable and reliable correlations than trying to determine them directly from the covariance matrices of forecast errors at different levels.

The correlation model is

$$C_{ij} = e^{-a\left|x_i - x_j\right|^{\alpha}} \quad \text{with } 1 \leq \alpha \leq 2 \qquad (4)$$

where a is an arbitrary tuning constant (not used by Cats).

In order to get finite temperature variances (entailing taking the derivative of (4)) and make the computations efficient, the exponent α should be 2 instead of 1.6 which is used for the present operational correlation.

The values of x_i define the coordinate transformation from the pressure levels and can be found by setting j=i-1 in (4).

$$x_1 = 0 \qquad (5)$$

$$x_i - x_{i-1} = \left(-\frac{1}{a} \ln C_{i,i-1}\right)^{\frac{1}{2}} \qquad (6)$$

$C_{i,i-1}$ is calculated by using the thickness and height error variances in (3). The model (4) is then used to define correlations at all the other levels.

The procedure for constructing the continuous representation of forecast errors and correlations from observed errors and correlations is as follows:

(a) Fit the height forecast errors at pressure levels to a sixth degree polynomial.

(b) Take the calculated height error correlations between adjacent pressure levels given by (3)* and compute the x_i's for all levels from (5) and (6) with a first guess value of parameter a. These are then fitted to a sixth order polynomial in ln(p) to give a continuous representation.

(c) With the correlation model (4) and the functional expression for x from (b), find a new value of a which fits the two off-diagonal correlations, i.e. $C_{i,i+1}$ and $C_{i,i+2}$, as well as $C_{i,i-1}$ and $C_{i,i-2}$ where applicable.

An example of the correlation function thus obtained is shown in Fig. 2. The 'observed' discrete correlations to be fitted were however already constructed using Cats' procedure, (though with an exponent of 1.6 instead of 2), so the fitting is relatively trivial although it also depends on how well the error variances and transformations can be fitted. The winds (u and v vertical statistics are identical) are fitted using the analogous procedure to that used for the heights.

---

*Note: The z_i's and z_j's in (3) are the fitted ones according to a) but the thickness variance is the observed one. This results in better thickness variances implied from the fitted functions but makes this procedure unsuitable for very thin thicknesses (i.e. temperatures).

2.3 The revised vertical correlation model

The correlation model with $\alpha=2$ in (4) was used for a 10-day assimilation and worked satisfactorily except for a rather large negative height bias in the tropical stratosphere. This was found to be caused by the form of the chosen correlation function. Even though z-z correlations are fitted relatively well, the Δz-Δz correlations did not adequately reproduce the discrete ones (Table 1); they were too broad and too large, resulting in a poor fit to thickness data in the stratosphere.

Instead of (4) a different functional expression was used which resembles (4) with $\alpha=1.6$ much better

$$C_{ij} = \frac{1}{e-1}\, e^{\left[\frac{a}{a+(x_i-x_j)^2}\right]} - \frac{1}{e-1} \tag{7}$$

a is again a tuning constant and the expression for the transformation x_i is:

$$x_i - x_{i-1} = \sqrt{\frac{a}{\ln\left(\frac{C_{ij}+b}{b}\right)} - a} \quad \text{where } b = \frac{1}{e-1} \tag{8}$$

The use of this formulation cures the problem of the stratospheric biases (see Sect.5). This fit to z-z and v-v correlations is similar or better than before except for the tropical v-v correlation at the highest levels. There the sharp transition between broad correlations in the stratosphere and extremely narrow correlations around the tropopause gives a poor fit for the transformation, and consequently for the correlations. The implied z-Δz and Δz-Δz correlations now reproduce the discrete ones much better, and in particular the Δz-Δz correlation is much narrower and smaller (see Table 1).

a)

100													
36	100												
-5	21	100											
-11	-6	28	100										
-7	-7	3	37	100									
-4	-4	0	14	48	100								
-6	-9	-15	-15	0	31	100							
-4	-6	-12	-16	-9	7	43	100						
-3	-6	-16	-29	-40	-39	-10	12	100					
0	-1	-4	-8	-11	-11	-3	3	39	100				
0	-1	-3	-9	-18	-25	-23	-26	-5	8	100			
0	0	-1	-3	-7	-11	-12	-15	-11	-15	30	100		
0	0	0	-1	-3	-6	-7	-10	-13	-24	-4	27	100	
0	0	0	0	-1	-3	-4	-6	-10	-21	-16	-8	27	100

b)

99													
82	100												
18	55	99											
-19	-14	50	100										
-16	-29	-14	59	100									
-7	-20	-38	-3	69	100								
-3	-11	-35	-36	18	78	100							
-1	-4	-22	-44	-32	19	70	100						
0	0	-8	-27	-47	-41	-7	54	100					
0	0	-1	-8	-25	-43	-49	-26	49	100				
0	0	0	-1	-8	-20	-35	-46	-20	58	99			
0	0	0	0	-1	-5	-12	-26	-42	-16	51	100		
0	0	0	0	0	0	-2	-7	-23	-37	-14	60	100	
0	0	0	0	0	0	0	-1	-9	-26	-40	-15	51	100

c)

99													
67	99												
3	35	99											
-7	-13	33	99										
-4	-12	-15	62	99									
-2	-7	-20	-12	54	99								
-2	-5	-17	-24	2	67	99							
-1	-4	-12	-23	-26	1	56	99						
0	-2	-8	-16	-27	-31	-17	37	99					
0	-1	-5	-9	-16	-23	-30	-27	30	99				
0	-1	-3	-5	-9	-13	-18	-26	-24	38	99			
0	0	-1	-2	-4	-6	-9	-14	-23	-20	31	99		
0	0	0	-1	-1	-2	-3	-5	-10	-17	-17	63	99	
0	0	0	-1	-1	-2	-3	-4	-7	-13	-21	-19	32	99

Table 1.
Tropical thickness-thickness correlations for 14 layers between the standard pressure levels and beginning with 1000-850 mb; in (a) are the discrete operational ones, in (b) continuous with eq.(4) and in (c) continuous with eq.(7).

3. THE COMPUTATION OF THE MODEL LEVEL ANALYSIS INCREMENTS

3.1 The basic principle for model level analysis evaluation

The analysis increment at a gridpoint or level is as in (1); that is

$$a_k = (\underline{d}^T \cdot \underline{\underline{M}}^{-1})\ \underline{P}_k$$

The data vector $\underline{d}$ consists of all the normalized data used for one analysis box at standard pressure levels and the matrix $\underline{\underline{M}}$ consists of all forecast and observation error correlations between the positions of the data in $\underline{d}$. Then the vector $\underline{P}_k$ contains the forecast error correlations between all the data positions and the gridpoint k. The method described here will use data only at, or interpolated to, standard pressure levels just as in the current operational scheme. In other words $\underline{d}$ and $\underline{\underline{M}}$ are unchanged. The $\underline{P}$ vector, which correlates the data with the gridpoints, will however be changed. $\underline{P}$ is formulated as separable in the horizontal and vertical correlations

$$P_{i,k} = V(p_i, p_k)\ H(\underline{r}_i, \underline{r}_k) \qquad (9)$$

where p_i and p_k are the pressures of the gridpoint and the datum.

$H(\underline{r}_i, \underline{r}_k)$ is a continuous function of the positions of the gridpoint and the datum, and is used to give analysis increments (a_k) on a regular latitude-longitude grid. However, in the previous ECMWF operational system the vertical correlations $V(p_i, p_k)$ are only defined for a number of discrete pressures (standard levels) and the analysis increments are given only for those pressure levels.

If instead of using correlations at discrete levels a continuous functional expression for $V(p_i, p_k)$ is used, then the correlation model (9) will enable us to get analysis increments at arbitrary positions <u>both</u> in the horizontal and vertical.

The $\underline{P}_k$ vector will now consist of correlations between data positions and gridpoints in a 1.875° latitude-longitude grid (as used at present), but at hybrid model levels instead of at standard pressure levels.

3.2 The procedure for calculating the increments

The procedure chosen here utilizes the data coefficients ($\underline{d}^T \cdot \underline{\underline{M}}^{-1}$) twice. In the first scan through these data coefficients the standard pressure level increments are evaluated, plus the height increment at the first guess surface pressure. The surface pressure is then incremented hydrostatically by

$$\frac{g}{R}\frac{p_s}{T_{2m}}\frac{\Delta z}{[1+(\frac{1}{\varepsilon}-1)q_{2m}]} \qquad (10)$$

with the aid of the first guess of temperature (T_{2m}) and specific humidity (q_{2m}) at 2 m.

By using the analysed surface pressure (p_s), the pressures at hybrid levels can be calculated. Then heights and wind increments are calculated at these pressures. The heights are evaluated at 'half' levels defined as

$$A_{k+\frac{1}{2}} + B_{k+\frac{1}{2}} \quad p_s$$

except for the highest level ($A_{\frac{1}{2}}, B_{\frac{1}{2}}=0$) where the pressure is set to

$$p_{\frac{1}{2}} = \exp\left\{2\ln[0.5(A_{\frac{1}{2}}+A_{1\frac{1}{2}})] - \ln[A_{1\frac{1}{2}}]\right\} \qquad (11)$$

The wind levels are defined as

$$p_k = 0.5\left[A_{k+\frac{1}{2}} + B_{k+\frac{1}{2}}p_s) + (A_{k+1\frac{1}{2}} + B_{k+1\frac{1}{2}}\, p_s)\right] \qquad (12)$$

i.e. the full model levels.

There is, however, a complication because the analysis is often done in three vertical slabs and the model levels will intersect slab boundaries. This means that a model level in one analysis box can either be in the bottom slab or in the middle slab or in between the two slabs, all in the same analysis box depending on the orography. To accommodate this, each gridpoint has weights (being 0 or 1) assigned to it according to which slab it is in. Gridpoints in between two slabs (or the one closest if there are none in the layer) are overlapped. The analysis evaluations from the two slabs are simply averaged.

The decision about which slab a gridpoint belongs to is taken by using the first guess surface pressure, so that the p_s evaluation will be consistent with the model level height and wind evaluations.

The heights calculated at model 'half' levels are used to calculate virtual temperature increments at full model levels.

$$\Delta T_{v_k} = - \frac{g}{R \ln r} (\Delta z_{k+1\frac{1}{2}} - \Delta z_{k+\frac{1}{2}}) \text{ where } r = \frac{(A_{k+1\frac{1}{2}} + B_{k+1\frac{1}{2}}\, p_s)}{(A_{k+\frac{1}{2}} + B_{k+\frac{1}{2}}\, p_s)} \qquad (13)$$

However when the two height increments stem from different slabs which contain different data an inconsistency may arise. This can be quite severe and can cause very unrealistic temperatures, especially in thin model layers near the lower boundary. To overcome this problem an overlapping scheme has been designed and is described in Sect. 4.

The hybrid level increments of virtual temperature, winds and surface pressure are spectrally fitted and interpolated to the model's Gaussian grid. Then these fields are added to the first guess hybrid level Gaussian fields (after the first guess temperature has been converted to virtual temperature). The humidity analysis is still evaluated on pressure levels mainly because it is intrinsically an analysis of precipitable water content in layers between standard pressure levels.

The humidity increments, as well as the stratospheric corrections applied to the first guess pressure level fields, are transformed to the Gaussian grid and then interpolated vertically to hybrid levels in the way described in the ECMWF Data Assimilation Scientific Documentation. These fields are then added to the hybrid level analysis fields of virtual temperature, winds and surface pressure, and the first guess specific humidity. Also the virtual temperature is converted back to the real temperature with the aid of analysed humidity. These fields then undergo a spectral transformation for use in the initialisation and forecast.

3.3 Computational overheads

It should also be mentioned that there are some extra costs in terms of computing time when doing the model level evaluation. Firstly the vertical correlations between pressure level data and hybrid model level pressures must be calculated for each gridpoint. However, since the data at present are only given at standard pressure levels, these correlations can be precalculated and redundant computations for data at the same level are avoided. Secondly the evaluation is done twice through all data and gridpoints, both at pressure levels and at model levels. This is the major extra overhead involved but part of it is unavoidable because the first scan through the data is necessary to analyse the surface pressure which in turn will determine the hybrid level pressures. The pressure level analysis is now in principle redundant but it is still used to provide first guess thicknesses for the humidity analysis and to provide a persistence forecast for the stratospheric correction scheme in the next analysis cycle. It may also be desirable to keep the pressure level analysis as the most accurate estimate of the standard pressure level fields for dissemination and diagnostic purposes (e.g. observation statistics).

The cost incurred by using this scheme was initially about 4 minutes extra CPU-time on the CRAY-XMP at the main analysis hours 00Z and 12Z. That is to be compared with a total CPU-time of about 550 seconds and an elapsed time of around 15 minutes for the operational scheme. Now, since the gridpoint evaluation part of the analysis became the dominant part, further optimizations were sought. The same horizontal correlations are used in both scans,and precalculating these correlations and saving them for the model level scan gave a substantial saving in time. The extra time consumption was thee reduced to 120-140 seconds CPU-time for the various analysis cycles.

4. THE VERTICAL OVERLAP CORRECTION

The ECMWF analysis system is designed to use all the data in boxes with horizontal sides of about 660 km in addition to some data from neighbouring boxes. The maximum number of data to be used is set to 191. When this number is exceeded there is a vertical division of the box into 3 slabs: 10-100 mb, 150-700 mb and 850-1000 mb. At 12Z and 00Z such a division occurs in about 70% of the boxes, but this is reduced to 20-25% in the intermediate hours. Due to the fact that different data have been used in the 3 slabs, the analysis at neighbouring levels can be quite inconsistent in some cases. In order to introduce some smoothing the analysis evaluations from adjacent slabs are overlapped and averaged at 150 mb and 700 mb. But still the temperature increments in the overlapped layer can be spurious. In the previous operational analysis this could cause a problem over oceans when using sea level pressure from a SHIP and thickness data from SATEM in the lowest slab, and only thickness data in the higher slabs. When the evaluation of the analysis at model levels was tried, very unphysical temperature increments resulted when the lowest σ-layer intersected the slab boundary over high terrain. Therefore an overlap correction scheme for heights had to be designed for both model level evaluation and pressure level evaluation.

The principle of the scheme is simple. The analysis should produce heights that are such that the resulting thicknesses are always analysed thicknesses or a linear combination of analysed thicknesses from adjacent slabs. The analysed values are $a_k = \underline{C}^T \underline{P}_k$ where $\underline{C}$ is the data coefficients $(\underline{d}^T \cdot \underline{\underline{M}}^{-1})$ in (1), resulting from one set of data and the inversion of its correlation matrix ($\underline{\underline{M}}$).

To quantify the correction scheme let us consider the general case where more than one level may be overlapped from adjacent analysis slabs.

The height analysis from slab I is evaluated at levels 1,2,3...N-1,N. The analysis in slab II is then evaluated at levels N-k, N-k+1, N-k+2...M-1,M, where N and M are the upper slab boundaries and $k \geq 0$ determines the degree of overlapping between the two slabs.

If we first consider the case when $k \geq 1$ (i.e. two or more levels overlap) the height analysis at level N-k is $0.5(z^{II}_{N-k} + z^{I}_{N-k})$ where z^{I} and z^{II} denote the height analyses from the different slabs (i.e. using different data). The thickness of the layer below is then

$$\Delta z_{N-k} = 0.5\ (z^{II}_{N-k} + z^{I}_{N-k}) - z^{I}_{N-k-1} \qquad (14)$$

This thickness is clearly spurious since different data have been used to analyse z^{II}_{N-k} and z^{I}_{N-k}. The aim of the overlapping correction scheme is to replace the thickness in (14) by an analysed thickness, namely

$$\Delta z'_{N-k} = z^{I}_{N-k} - z^{I}_{N-k-1} \qquad (15)$$

This is accomplished by adding a correction term at level N-k and all levels above

$$\Delta z'_{N-k} - \Delta z_{N-k} = 0.5(z^{I}_{N-k} - z^{II}_{N-k}) \qquad (16)$$

In other words the spurious thickness $(\Delta z'_{n-k} - \Delta z_{N-k})$ is removed.

Now another problem arises in the layer above the lower slab (I). There the evaluated thickness would be

$$\Delta z_{N+1} = z^{II}_{N+1} - 0.5(z^{I}_{N} + z^{II}_{N}) \tag{17}$$

The correct analysed thickness should be

$$\Delta z'_{N+1} = z^{II}_{N+1} - z^{II}_{N} \tag{18}$$

and the correction term to be added to the height analysis at level N+1 and all levels above is

$$\Delta z'_{N+1} - \Delta z_{n+1} = 0.5(z^{I}_{N} - z^{II}_{N}) \tag{19}$$

Thus for such an overlap layer there are two corrections to the height analyses - one at level N-k and above [using (16)], and one at level N+1 and upwards [using (19)].

In the case of k=0, i.e. only one level is overlapped, the only difference is that the two corrections give the same value and it is added once at level N and twice at all levels above.

5. EXPERIMENTS AND RESULTS

5.1 Experiment 1

A 10-day assimilation using the continuous representation of error variances and correlations, with the formulation given in (4), was run for the same period as the one used for testing the comprehensive analysis changes which became operational in May 1984 (see Appendix I for details of the experiments). This baseline assimilation contained all those changes but did not include the diurnal cycle which also was introduced operationally in May 1984.

The negative height biases in the higher troposphere are reduced throughout the assimilation (see Fig.3), especially for first guess and initialisation (note - the analysis in Fig. 3 is the one evaluated at pressure levels which is not used for the initialisation and forecast in the experiments).

The biases are markedly reduced in the tropics. However in the stratosphere the negative height bias is increased (especially in the tropics). Also the analysed thickness bias and RMS-fit deteriorated and is now closer to the first guess (see Fig. 4), i.e. the observed thicknesses between adjacent levels are not fitted as well.

The first guess and initialised heights and winds have a slightly better RMS-fit to TEMP data in the tropics. The analyses are otherwise generally very similar to the baseline. The first guess winds are somewhat stronger in this assimilation, probably caused by the direct evaluation. The equatorial easterlies at 30 mb (QBO) are especially accentuated in this case.

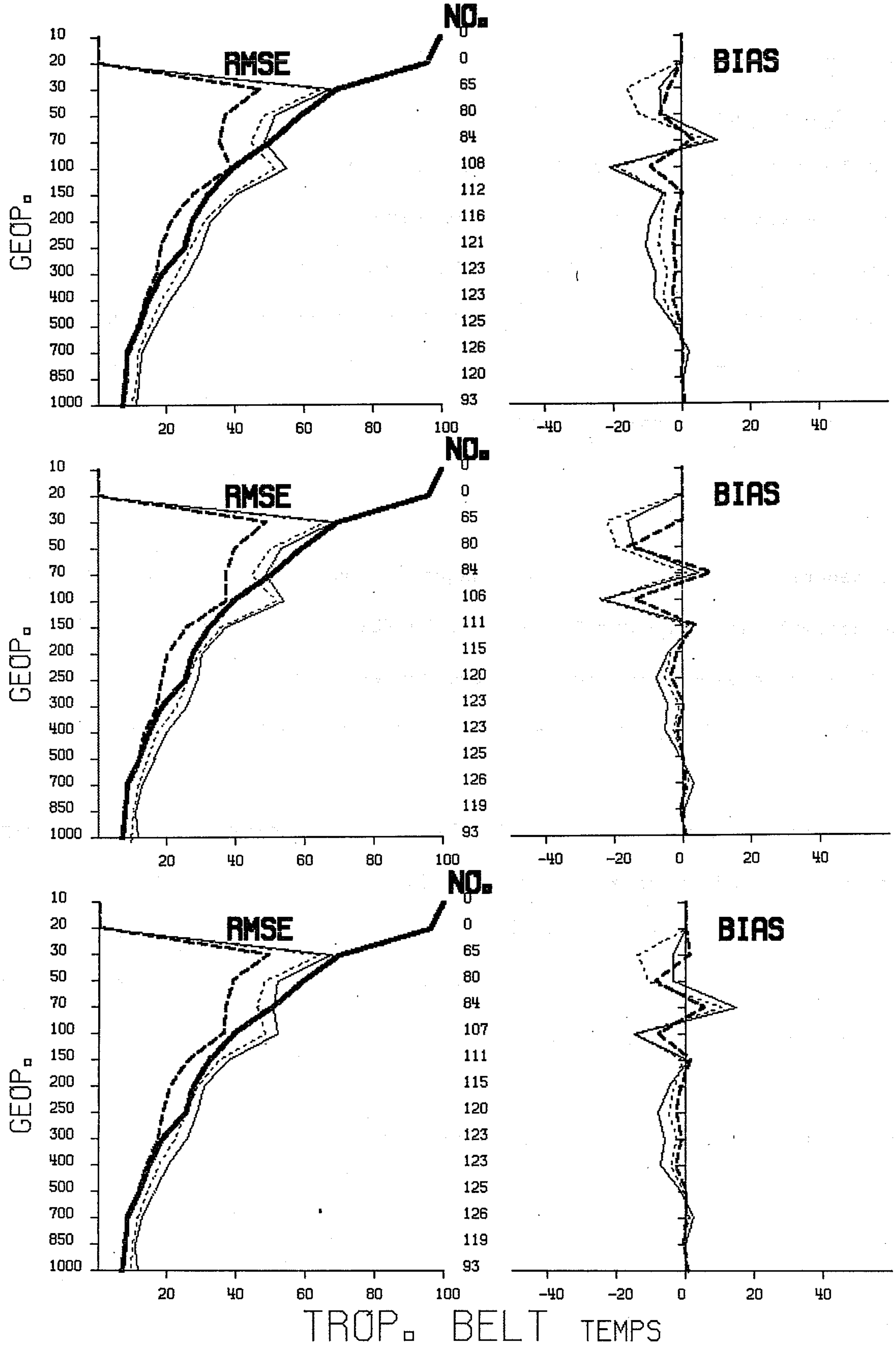


Fig. 3 RMS deviations and biases between TEMP heights (m) and first guess (full line), analysis (thick dashed) and initialised height (thin dashed line) at 12Z 1983-11-28.

Top-baseline, middle-first experiment using the correlation function given by eq.(4), bottom-second experiment using eq.(7).

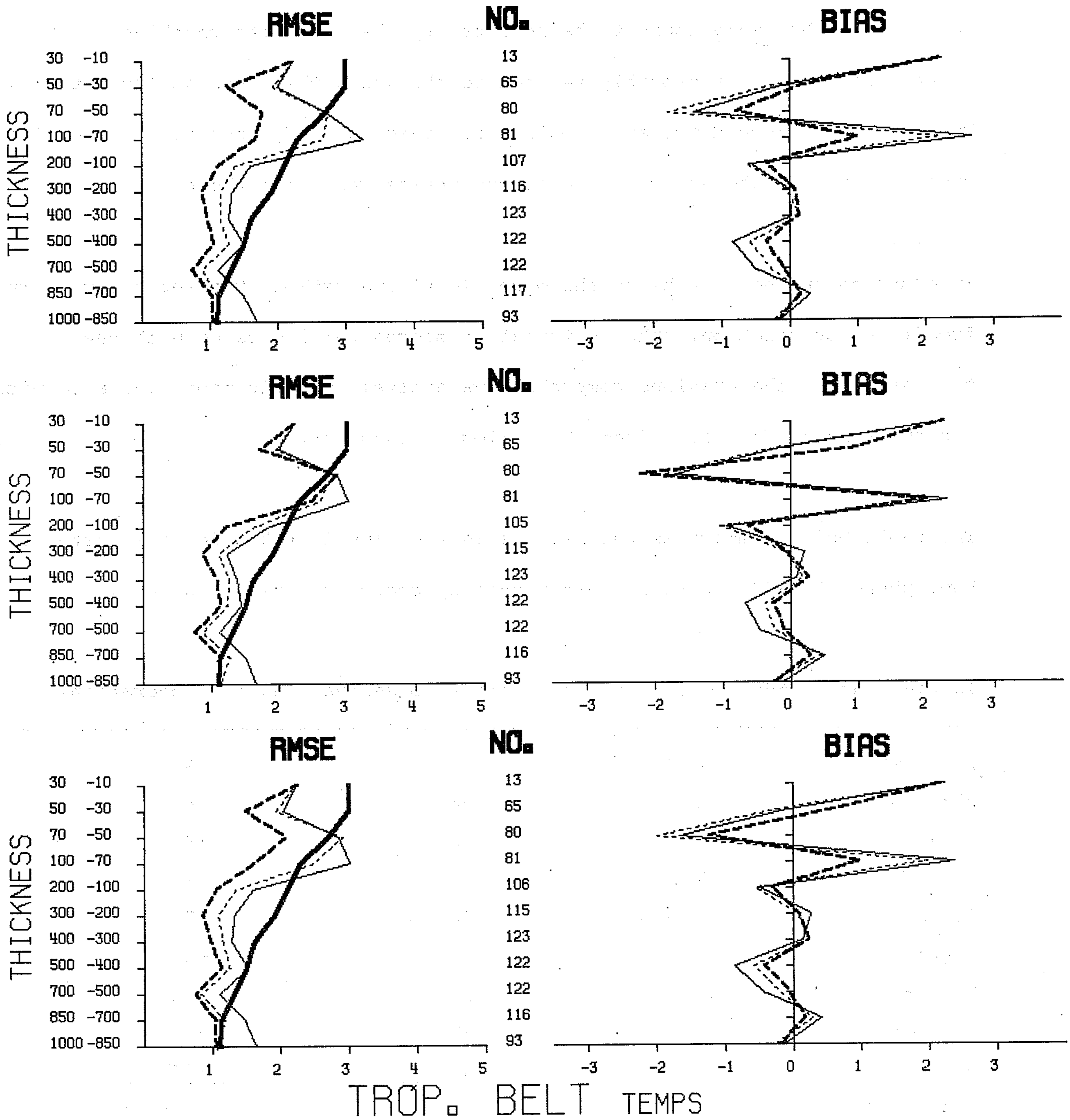


Fig. 4 RMS deviations and biases of thicknesses converted to mean temperature (K) also at 12Z 1983-11-28. The same layout as in Fig. 3.

Top-baseline, middle-first experiment using the correlation function given by eq.(4), bottom-second experiment using eq.(7).

A more worrying aspect in the assimilation was that at certain 12Z analyses there were up to 10-12 more TEMP heights rejected. When these were scrutinized they were found to be rejected by the O.I. data checking in the stratosphere and thus probably related to the worse thickness analysis there. The fact that these detrimental thickness biases already appeared in the first cycle pointed to the structure functions themselves being the culprits.

In order to further benchmark the model level evaluation, 9 forecasts were run from this assimilation. The verification scores were compared with the forecasts from the previous comprehensive analysis modifications and also with the operational forecasts from the period in question.

The table below summarises the results as the time it takes for the northern hemisphere 1000-200 mb total height anomaly scores to fall below 60%.

Initial date	60% predictability	Gain vs baseline	Gain vs operations
26/11/83	5d 1h	+8h	-3h
27/11/83	5d 8h	-½h	-4h
28/11/83	4d 15h	-5h	+4h
29/11/83	4d 2h	-2h	-3h
30/11/83	5d 10h	-4h	+7h
1/12/83	5d 5h	-4h	-2h
2/12/83	5d 6h	+½h	+3h
3/12/83	4d 18h	+8h	+15h
4/12/83	4d 4h	+4h	-11h
Mean:		+½(±1)h	+½(±1)h

The scores indicate a slight improvement, but it is not very significant. However they do show that even though the structure functions in this experiment were not entirely satisfactory they did not cause a deterioration in the forecasts and the direct evaluation give at least as good forecasts as the baseline.

5.2 Experiment 2

The correlation model in (7) was found to give better Δz-Δz correlations and a second assimilation was started using this new formulation. Also the observed thickness errors were used for the transformation described in Sect.2.2 rather than a modelled one which was used before in an iterative way. Now the extra negative height bias in the stratosphere was removed and the thickness analyses were almost as well fitted to data as the baseline (Figs.3 and 4). As a consequence the rejection rate was now almost the same as in the baseline (0-2 fewer at 00Z and 0-5 more rejected at the 12Z analysis cycles).

This second assimilation was interrupted after 4½ days due to tape problems during the holiday period when it was intended to run. The results seemed, however, conclusive and, since a more recent summer case was desired, the assimilation was not pursued any further. Only three 10-day forecasts were made. Their results are summarized below and show improvement, in two cases significantly so, compared to the first experiment.

Initial date	60% predictability	Gain vs baseline	Gain vs operations
26/11/83	5d 2h	+9h	-2h
27/11/83	5d 10h	+2h	-2h
28/11/83	4d 12h	-8h	+1h

It is encouraging that an experiment with structure functions which look better not only assimilate the data better but also give a positive signal in the forecast scores. Also there is a tendency to make the scores less variable compared to the then operational forecasts using the old data assimilation system.

5.3 Experiment 3

The last case to be tested was a week in August 1984 (see Appendix I) and the assimilating model now included the diurnal cycle (it became operational in May 1984). The analysis used the correlation model (7) as in the previous experiment.

The analyses again agree well with the operational one (which now included the comprehensive data assimilation changes). Rejections were almost identical - only at some 12Z cycles were there a few more TEMP data rejected. The negative height biases, especially in the first guess and the initialised field, were reduced in the troposphere. This is shown for the tropical belt (20°N-20°S) in Fig. 5 which includes all used TEMP heights for all the 12Z assimilation cycles during the week. The RMS-fit of height data is only marginally better in the troposphere. The winds data from TEMP and PILOTS have a slightly improved RMS-fit in the tropical troposphere but it is only 0.1-0.2 m/s better.

The results from the 6 forecasts run for this period are shown below and they indicate a neutral impact on the scores.

Initial date	60% predictability	Gain vs operations
24/08/84	6d 12h	+4h
22/08/84	6d 19h	±0h
23/08/84	5d 19h	-1h
24/08/84	5d 1h	-1h
25/08/84	4d 15h	-1h
26/08/84	4d 20h	-1h
Mean:		±0(±1)h

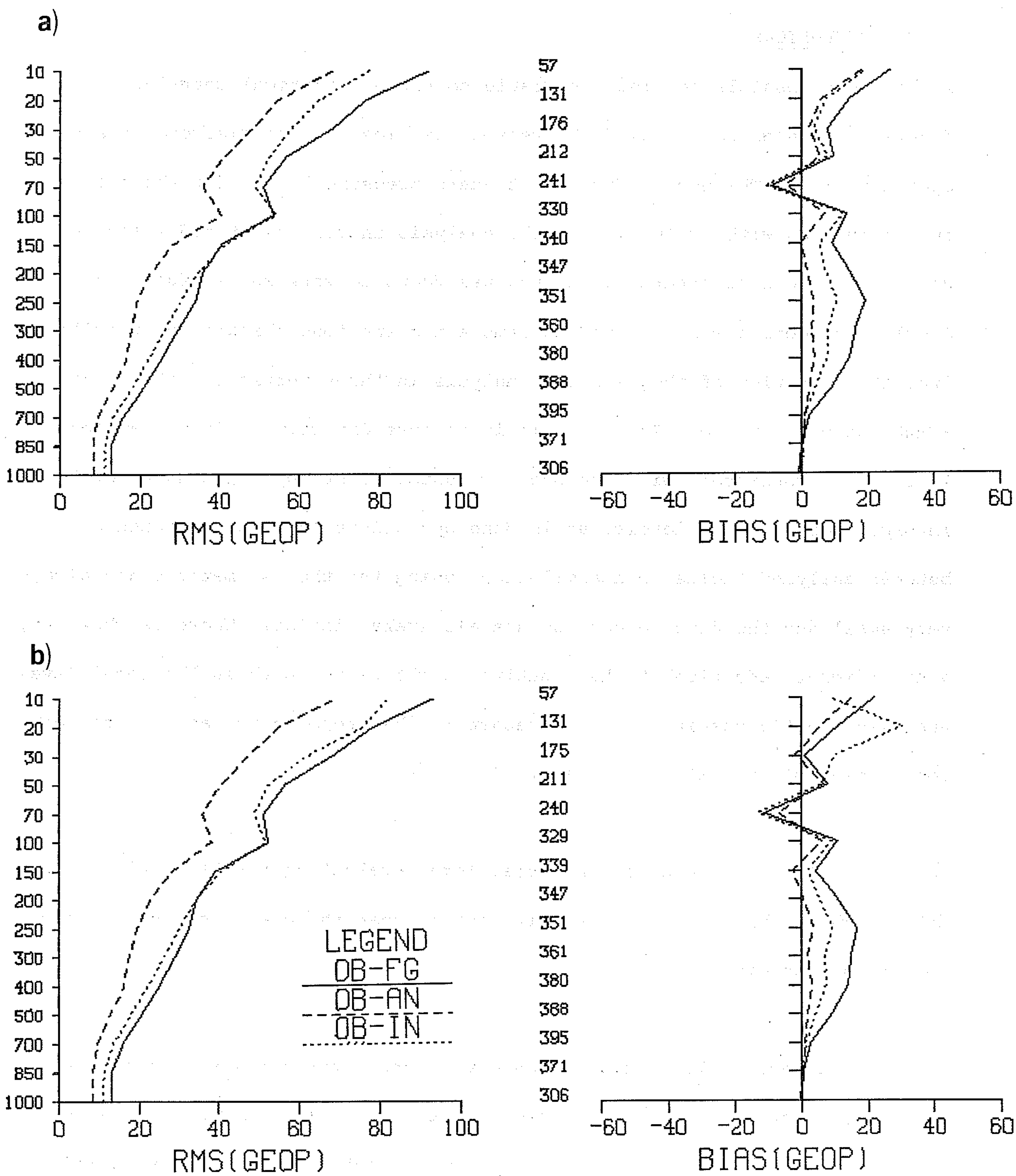


Fig. 5 Composite root mean square (RMS) fit and bias between TEMP heights and first guess, initialisation and analysis. The operational statistics for 12Z 19.8.84-26.8.84 are given in (a), whilst the 12 Z results for experiment 3 (using eq.(7)) are shown in (b). Note that in this figure the bias is given by (obs.-field) whereas in Figs.3 and 4 the bias is (field-obs). The initialisation statistics are not computed above 50mb because of the operational stratospheric correction.

6. CONCLUSIONS

It has been possible to find a suitable continuous vertical correlation function which gives correlations reasonably close to the discrete ones used operationally. The use of data at standard pressure levels for the analysis, in combination with evaluation of the analysis increments at model levels using continuous structure functions, was found to work successfully once an overlap correction was designed to remove the spurious thicknesses resulting from the partition of the data and analysis in three vertical slabs in the ECMWF analysis system. The analysis increments directly evaluated at model levels were consistent with the ones evaluated at standard pressure levels and interpolated to model levels, as is done operationally. The differences between analysed fields in assimilations using the the two methods are always very small and the data rejections are also very similar. There is, however, a significant reduction of the negative height biases both in the model level analysis and the ensuing 6-hour forecast in the troposphere, as well as some improvement of the RMS-fit to data.

As for the 10-day forecasts, the model level evaluation resulted in forecasts only slightly (but not significantly) better than the ones from the current operational scheme.

Since the direct analysis evaluation gives better first guess forecasts and some improvement in the 10-day forecasts, albeit small, it was considered desirable to implement it operationally (implemented on 13 November 1984). Also the continuous structure functions will form a basis for the new ECMWF analysis system and it would be valuable to gain further experience by using them within the existing system.

REFERENCES

Cats, G., 1982: Constructions of a vertical correlation matrix. ECMWF Res.Dept.Memo R2302/672.

ECMWF Data Assimilation Scientific Documentation. ECMWF 1983.

Hollingsworth, A., 1982: Analysis in σ-coordinates. ECMWF Res.Dept.Memo.R2294/777.

APPENDIX I - OVERVIEW OF THE DATA ASSIMILATIONS REFERENCED IN SECT.5

Experiment	Time period	Properties of the experiment
Operational assimilation	12Z 24/11/83- 12Z 04/12/83	-
Baseline assimilation	12Z 24/11/83- 12Z 04/12/83	Comprehensive data assimilation changes (D.A.) as introduced in May 1984 (e.g new structure functuons, revised data selection, rejection criteria and error growth and overlap correction).
Experiment 1	12Z 24/11/83- 12Z 04/12/83	Comprehensive D.A. changes, as above but model level evaluation using correlation function as Eqn.4. Fitted heights and thicknesses were used for the transformation.
Experiment 2	12Z 24/11/83- 00Z 29/11/83	As above but using Eqn.7 and observed thickness errors for transformation.
Operational assimilation	00Z 12/08/84- 12Z 26/08/84	Comprehensive D.A. changes and diurnal cycle
Experiment 3	12Z 19/08/84- 12Z 26/08/84	As above but model level evaluation using Eqn. 7 and observed thickness errors.